

Responding to uncertainty

Public controversies that involve scientific uncertainty can be influenced by mavericks. Open confrontation and analysis serves the public better than excommunication.

In May 2002, the Science Media Centre (SMC), a UK organization dedicated to providing journalists with access to scientists, conducted a closed seminar in which government officials, reporters, researchers and others reviewed a calamity of communication and of public response to science. The flashpoint had been some ill-judged remarks made at a press conference in late 2001, suggesting that the triple-vaccine regimen supplied to the UK population against measles, mumps and rubella might be associated with autism. What followed were campaigns by the government to reassure parents, and by parents for separate vaccines on demand and against alleged conspiracies by the scientific establishment. As a result, many children were not vaccinated, and deaths resulted that could have been avoided.

The seminar concluded on a dilemma that still faces any government faced by a crisis bedevilled with scientific uncertainties, such as today's threat of avian flu. Is it better for authorities simply to reassure, resisting discussion of uncertainties in the expectation that the public would be paralysed or panic-stricken by the lack of clarity? Or should leaders assume a degree of maturity on the part of the media and public, and represent the state of the science, risks and all?

Nature would always urge the latter. But the mass media find it hard to handle scientific uncertainty — and all the more so when vocal scientists promote minority views that have their own appeal to segments of the public, and which gain profile because of media obligations to provide balance. All the more onus on other scientists, then, to help journalists with analyses for public consumption.

Learned societies have a particular responsibility in such circumstances. They should not wait to be called on by a government to provide an analysis. Rather, where the stakes for public interests are high, they should promptly convene a working group charged with delivering a statement of the state of relevant evidence, within a few days if necessary. Above all, they should provide the media with succinct statements and readily reproducible graphics that clearly illustrate not only the conclusions drawn by the working group but also the state of the uncertainties. And where there is

a maverick voice, that extreme perspective needs to be exhibited as such. It should be presented in the context of the range of scientific judgement: not dismissed by assertion, but discussed and visualized against a background of expert opinion and the conclusions of studies.

Some societies have been known to lobby the media publicly or discreetly to try to discourage them from allowing minority voices to be heard. Nothing could be more counter-productive. Even if a high-profile scientist is judged by peers to be lacking credibility, the media will rightly be provoked by attempts at censorship, which fuel allegations of a conspiracy, adding perceived weight to maverick claims. It is better to attack such claims explicitly on a scientific basis.

Such problems arise in any scientific country. The SMC has made a particular contribution to mitigating them in Britain. The brainchild of Susan Greenfield, the director of the Royal Institution of Great Britain, which hosts it, its success can be credited above all to the robust

"The UK Science Media Centre provides quotes from experts in immediate response to breaking stories, and in-depth briefings for longer-running controversies."

leadership of its director Fiona Fox. It provides quotes from experts in immediate response to breaking stories, and in-depth briefings for longer-running controversies. It tutors scientists in communicating complexities such as risk with respectable but effective soundbites. In all of this it acts independently, on behalf of both journalists and scientists — but it ultimately serves the media. Other countries are beginning to take note of the SMC, and Australia is set to clone it.

The media cannot always be trusted — sometimes for lack of resources and knowledge, or for rank editorial bias. But whether scientists take initiatives themselves through campaigning and blogging, or learned societies get sharp, or intermediaries such as the SMC are established, journalists and the public need to be treated as the sophisticated recipients of prompt and well tailored information that most of them are.

Still not deterred

Universities should back researchers determined to stand up for animal research in the face of terrorism.

Around 18 months ago, under the title "Defeated but not deterred", this magazine published an editorial on animal-rights activism in Britain. A campaign of violence and vandalism had just forced the University of Cambridge to cancel plans for a primate research laboratory. *Nature* suggested then that the

victory would not necessarily be repeated. We wish we had been right, but that optimism now looks premature.

The extremists have now secured another success. Darley Oaks is a small family farm in the Midlands. The owners have been breeding guinea pigs for medical research for more than a decade. For the last six of those years, they have suffered a campaign of arson attacks and death threats, and have seen letters of abuse sent to people connected with their business. The campaign culminated in macabre fashion last October, when the remains of a family relative were dug up and removed by protestors demanding that the business close. Last week, the family decided to comply.



It was the second notable victory for the animal-rights extremists since the Cambridge decision. Work on a new animal house at the University of Oxford has been stalled for over a year, after threats and vandalism forced contractors to pull out. The increased support for animal research seen in the media and in public opinion polls, and highlighted in our editorial, has continued during that period. But the activists have become more focused and strategic — and more successful.

Analysis of the new tactics suggests that the fight against the extremists may take longer to win than anticipated. Activists have been concentrating on a small number of secondary targets, such as financial firms that act for animal testing laboratories. Attacks to individuals' homes are common. It is impossible for the police to defend such a wide range of targets and, as a consequence, many companies have ended associations with animal researchers and breeders. And it is not just British researchers who should be concerned: similar tactics are becoming increasingly common in the United States, where the FBI is investigating more than 100 cases, many involving arson.

Yet there is plenty to suggest that the extremists' actions will only prolong their fight, not allow them to win it. New UK legislation, aimed at punishing protestors who set out to cause economic damage to companies, came into force this summer. Perhaps because of publicity surrounding the new offences, attacks on company and private property connected with animal research fell sharply even before the new powers became law.

In the longer term, police must infiltrate the activists as they

would any other extremist organization. It is a welcome sign that the UK government has made animal activists a major focus for the National Extremism Tactical Co-ordination Unit, a policing body established in March 2004.

Action from the research community is the final front in the fight. Here the news is encouraging, although a great deal remains to be done. A decade ago, the Research Defence Society (now known as the RDS), which supports the medical use of animals, could call on just two researchers when asked to supply scientists for media interviews. That number now stands at 25. This is still far too few, given that thousands of researchers in Britain use animals in their work. But it is a start. It is also good to see that more than 500 prominent scientists have put their names to an RDS statement released last week, the Declaration on Animals in Medical Research, that backs animal research.

Perhaps the only group not pulling its weight is the university sector. Some universities encourage researchers to speak out, but they are in the minority. Universities have just as big a stake in animal research as any other science organization. This reluctance to talk to the public about their research merely plays into the hands of the extremists, who would be delighted to see scientists stay silent. ■

"More than 500 prominent scientists have put their names to a declaration released last week that backs animal research."

Three cheers

This issue of *Nature* includes several reasons for editorial celebration.

This is, above all, the issue of the chimp. It is not just the formal publication of the genome and related analyses, which would be ample cause for celebration. By a happy coincidence, the first fossil chimp has been found. This cornucopia of new science is introduced on page 47. Elsewhere in the issue, and in our free-access news service (www.nature.com/news), readers can find a range of discussions of comparative genomics, threats to chimp populations, a timeline of chimp research, and an introduction to famous chimps.

Alongside all this progress in chimp biology there are some provocative but essential questions raised about the future of conservation and of research, in Commentaries by Pascal Gagneux and colleagues (on page 27) and by John VandeBerg and Stuart Zola (on page 30).

What else to celebrate? In its role as a journal, *Nature* has historically focused on the formal reporting of research results and, like all journals, has ignored the process of discovery. In one small way we breached that barrier a few years ago by inviting authors to spell out their individual contributions at the end of the paper. After a slow start, more and more authors are responding to this invitation.

This week we go one step further, with the launch of our weekly

'Authors' page (page xiii). This is intended to show the joys, dead-ends and happenstance that lead to the 'eureka moment' of discovery. In 'Making the Paper', two authors tell how a hunt for some stone tools led them to a chimp tooth that has major implications for evolution. The 'Abstractions' column highlights individual contributions. This week, a computer programmer and database administrator recounts how she wrestled with an overabundance of chimp and human genome data and explains the difficulties in running a comparative analysis. Finally, 'Quantified' discusses some timely statistical data about *Nature* papers and provides some context behind the numbers.

So much for science fact. Regular readers will know that we have also been publishing a series of short science-fiction stories under the 'Futures' banner. The current series — a successor to that published at the turn of the Millennium — concentrates on developments that might occur within the next 50 years, roaming widely from the effects of climate change and genetic engineering to how the dead might be reunited as spam filters. The third cheer this week is for recognition by the European Science Fiction Society, which has bestowed its award of Best Publisher to the series editor Henry Gee and to *Nature*. ■

"The current series of 'Futures' concentrates on developments that might occur within the next 50 years, roaming widely from the effects of climate change to how the dead might be reunited as spam filters."

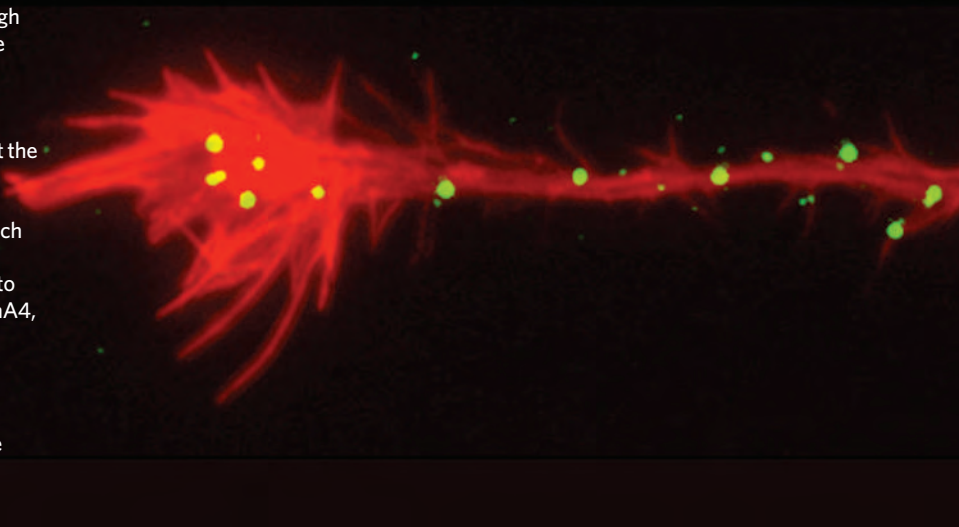
RESEARCH HIGHLIGHTS

Second signal*Neuron* **47**, 515–528 (2005)

One way that cells communicate is through receptor proteins called receptor tyrosine kinases. Typically, such receptors relay signals when a kinase enzyme that forms part of their structure is activated.

However, a team led by Rüdiger Klein at the Max Planck Institute of Neurobiology in Martinsried, Germany, has shown that signalling by receptors called Ephs — which guide axon growth — is not that simple.

The team genetically engineered mice to carry an altered form of an Eph called EphA4, whose kinase is permanently active. Unexpectedly, some aspects of neural development were still normal. The team suggests that clustering of receptors (pictured in green on the red growth cone of an axon) triggers signalling through an additional mechanism.



J. EGGER/K. DENNINGER/R. KLEIN

CANCER**Vessels take off***Genes Dev.* doi:10.1101/gad.1308805 (2005)

The tumour suppressor gene *PTEN* governs the formation of blood vessels by influencing vascular growth factors, suggests a study in mice. Animals lacking one copy of the gene showed unregulated expansion of vessels around tumours and faster cancer growth than controls. This is probably because the vessels feed and sustain the developing cancer.

Akira Suzuki of Japan's Akita University School of Medicine and his fellow authors also note that people suffering from hereditary disorders such as Cowden disease — which make them susceptible to tumours — carry mutated copies of *PTEN*.

dwarf star, and ions heated to 100,000 °C accreting on its surface. These characteristics are shared with bigger T Tauri stars, which are bright young things with less than twice the mass of the Sun.

ANIMAL BEHAVIOUR**Chicken little***Curr. Biol.* **15**, R620–R621 (2005)

Migrating birds are known to use the Earth's magnetic field to orient themselves, but researchers have struggled to condition birds to respond to fields in the lab. Now a team led by Rafael Freire from the University of New England in Armidale, Australia, has achieved this in an unlikely species: the domestic hen.

The researchers attribute their success to the use of a social stimulus, rather than the usual food, as a reward in their experiments.

Young domestic chickens were allowed to develop an attachment to a red ball and were then sent to find it. The direction of a local magnetic field was shown to influence the direction in which a chick started its search over a series of trials. This finding also reveals that the chicken's magnetic compass has survived thousands of years of domestication.

MICROBIOLOGY**Ancient disease***PLoS Pathogens* **1**, 5 (2005)

Tuberculosis may have affected early hominids, according to a genetic analysis that extends the pathogen's family tree.

Cristina Gutierrez from the Pasteur Institute in Paris and her colleagues isolated rare strains of tubercle bacilli from patients in east Africa. Genetic comparison with *Mycobacterium tuberculosis*, the agent responsible for most tuberculosis cases today, revealed that a common ancestor existed an estimated 3 million years ago. The comparison also showed that genetic recombination had occurred between the diverging strains.

The spread of a complex of strains resembling *M. tuberculosis* may have coincided with waves of human migration out of Africa, the researchers speculate.

ASTRONOMY**Growing dwarf***Astrophys. J.* **630**, L89–L91 (2005).

Brown dwarfs often seem like the runts of the astronomical litter, being too small to fuse hydrogen. But observations now confirm that they grow like proper stars.

Astronomers from the University of Delaware in Newark have used the Hubble Space Telescope to collect the first ultraviolet spectrum of a brown dwarf. Their study of object 2M 1207 reveals a cold cloud of molecular hydrogen surrounding the

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

NANOTECHNOLOGY**Drop by drop***Nature Mater.* doi:10.1038/1455 (2005)

Demonstrating exceptional chemical wizardry, a team of European researchers has harnessed light-powered molecular motion to drive a droplet of liquid up a sloping surface. The technique could be used in lab-on-a-chip devices.

The researchers coated the slope's surface with rotaxanes. These molecules consist of a chain threaded through a ring that shuttles up and down the structure. The team

designed the rotaxanes so that ultraviolet light would push the shuttle group to one end, hiding a fluoroalkane segment of the chain in the process. This makes the rotaxane layer more wettable where the light is shining. So, illuminating the front edge of a droplet drags it forward.

MATERIALS

Diamond geezers

Appl. Phys. Lett. **87**, 083106 (2005)

A diamond material that has been assembled from spherical carbon molecules (C_{60}) is denser and harder than the real thing (pictured), report Natalia

Dubrovinskaia from the University of Bayreuth in Germany and her colleagues.

The researchers squashed the fullerene molecules together at 2,200 °C using 20 gigapascals of pressure, which is equivalent to the weight of the Titanic pressing down on an area the size of a CD.

This left them with a translucent cylinder made from a jumble of diamond nanorods, each of which was less than 20 nanometres across. The material is 0.2–0.4% more dense than natural diamond, and tests suggest it could form longer-lasting coatings for precision drill bits.

ECOLOGY

Richer soil

Science **309**, 1387–1390 (2005)

The average gram of unpolluted soil contains a million different bacterial species, according to a reanalysis of existing data.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Previous estimates put the figure at 10,000.

The number of species is estimated by jumbling DNA from a sample and seeing how quickly matching sequences group together. Past calculations assumed that all species were equally common. To reach the higher figure, the team from Los Alamos National Laboratory in New Mexico allowed for variation. It also calculates that pollution with toxic metals wipes out 99.9% of bacterial species, despite an unchanged cell count.

EVOLUTION

Margin for error

Nature Genet. doi:10.1038/ng1621 (2005)

A classic problem in biology, concerning the origin of life, may benefit from a relaxed view.

Eigen's paradox points out that long genomes need complex enzymes to replicate themselves reliably — and questions whether the code for such enzymes could fit within

these genomes. Now Eörs Szathmáry of the Collegium Budapest in Hungary and his colleagues redefine what constitutes a 'reliable' copy by mapping the effect of mutations — or mistakes — in RNA molecules called ribozymes.

They show that a ribozyme's function, which is defined by its shape (or phenotype), can survive a relatively large number of mutations to its sequence (or genotype).

This suggests that long genomes could survive much higher error rates in copying than researchers had previously thought.

CELL BIOLOGY

Bubble wrap

Cell **122**, 605–617 (2005)

Within cells, many proteins are transported inside bubble-like lipid sacs called vesicles. These bud off the membrane layers of the endoplasmic reticulum, where such proteins are made. But how are the vesicles sculpted?

Researchers led by Randy Schekman at the University of California, Berkeley, have revealed that a protein called Sar1p initiates membrane curving by poking a helical arm into the membrane. If the arm is removed from the protein, budding never begins. Sar1p then interacts with other proteins that stabilize the budding vesicle and capture its cargo. And it seems that Sar1p is involved from start to finish, because buds formed in the presence of mutant Sar1p are unable to pinch off, remaining attached to the membrane layer.

JOURNAL CLUB

Craig Bina
Northwestern University,
Evanston, Illinois

A geophysicist revises his heat-flow notes by the light of a deep blue crystal.

In September 1999, after a long day at a conference, I was sharing a bottle of wine with Joe Smyth on the shore of Lake Maggiore in northern Italy.

He was rhapsodizing about the translucent, blue crystals he had

formed by squeezing olivine — a major constituent of Earth's mantle — to the pressures found deep beneath the surface.

A few weeks later I was back in the United States, teaching an undergraduate geophysics class. As I filled the blackboard with equations for heat transport in Earth's interior, I dutifully crossed out the radiative terms as negligible.

We have all known since the 1970s that rocks are opaque to radiation at the high pressures and temperatures of the mantle, so its

contribution to heat flow is deemed unimportant.

There matters rested, until I picked up the July issue of the *American Mineralogist*, and was reminded of Joe and those crossed-out terms.

Joe had measured the absorption spectra of his beautiful blue crystal — a hydrous ringwoodite — as it was squeezed between two diamonds to the pressure found 645 km down (H. Keppler and J. R. Smyth *Am. Mineral.* 1209–1212; 2005).

The surprising result? The

absorption of red and infrared wavelengths decreased as the pressure intensified. Thus, for realistic ringwoodite, radiative heat transfer is important.

This means that plates heat up more quickly than we thought as they subduct into the mantle, and it affects our modelling of mantle plumes such as those that produced the volcanic islands of Hawaii.

Not all the consequences are clear, but one thing is certain: I'm revising my lecture notes before classes begin.

NEWS

Advice on nuclear safety set for update in wake of floods

When Hurricane Katrina threatened the US coast, a nuclear power plant in the storm's path was shut down as a safety precaution. But, as last December's tsunami showed, there isn't always advance warning of floods.

Experts from 16 countries gathered in Kalpakkam, India, this week to discuss whether international safety standards for nuclear plants in flood-risk areas need to be modified. Kalpakkam is the site of India's prototype fast-breeder reactor, still under construction, which was partly flooded by 2004's Indian Ocean tsunami (see *Nature* 433, 675; 2005).

After that disaster, India's nuclear authorities hastily organized about a dozen national meetings to discuss lessons learned. This week's workshop, organized by the International Atomic Energy Agency (IAEA), was

the first time Indian scientists and officials shared their experiences with the international community.

India operates six nuclear plants, and another site is scheduled to start operating by 2008. "Practically all sites are at risk of being flooded," says Saurabh Verma, a researcher at the National Geophysical Research Institute in Hyderabad.

Outside India, nuclear facilities at risk include plants in Japan and the United States that sit along the tsunami-prone Pacific rim. But most of the world's more than 430 nuclear power plants need stronger protection against flooding, meeting organizers say. Plants are often located near coasts, where they use sea water to help cool the reactor.

Fewer than half of the countries that operate

nuclear power plants sent representatives to the meeting. Britain, Canada and Russia, for example, said they were unable to attend. But, based on the meeting's outcome, the IAEA will probably update its safety recommendations, which it last modified in 2003. National governments will then decide whether to adopt the voluntary standards.

The requirements include guidelines for the minimum distance of plants from the shore and the height and strength of protective walls. Post-tsunami surveys carried out in affected countries have yielded a wealth of information — about propagation of waves, run-up heights of water along different coasts, and the efficiency of flood warning systems — which safety experts can use to model the protection needed at given sites.

"From an engineering point of view, flood protection is easy," says Antonio Godoy, a senior safety officer with the IAEA. "The point is to come up with correct safety margins for specific sites, taking into account the local topography, flooding probability and inundation patterns. We need to ask ourselves, have we done enough or must we improve?"

In nuclear plants, flooding primarily affects the water intake and cooling systems. The greatest danger is that water could breach buildings that deal with plant safety, causing electrical short circuits or outright failures.

The worst flooding of a nuclear power plant in recent years occurred in December 1999 at the Blayais site in France. The flood, which was triggered by a combination of a storm surge, high waves and an 'ordinary' river flood, rated two out of seven on the International Nuclear Event Scale, but experts admit it came dangerously close to a genuine nuclear accident.

Power plants were also at risk during floods in Germany in 1997 and 2002. In the United States on 29 August, operators shut down the Waterford nuclear power plant near New Orleans as Hurricane Katrina approached. The US Nuclear Regulatory Commission sent additional staff to two other plants in the region.

"Tsunami risk is just one side of the problem," says Godoy. "A combination of events, such as extreme precipitation and a breach in a dyke, could be at least equally catastrophic." ■

Quirin Schiermeier

Additional reporting by Valeska Stephan

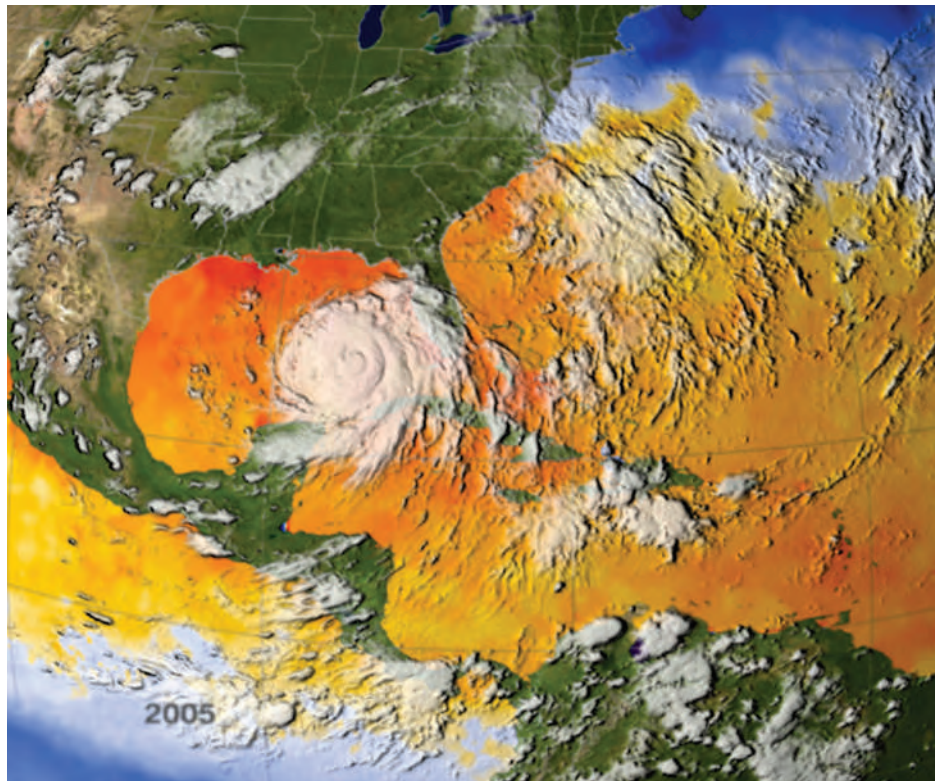
IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Alarms rang when foundations for a power plant in Kalpakkam, India, were flooded by the 2004 tsunami.



**TREES FAIL TO SUCK UP
EXCESS CARBON DIOXIDE**
Forests don't get a growth
spurt from greenhouse gas.
www.nature.com/news

UPI PHOTO/NASA-SVS/NEWS.COM



SNAPSHOT Sea heats up for hurricane season

Hurricane Katrina, which hit the US Gulf coast earlier this week, may have plenty of company this storm season. Warm sea-surface temperatures are fuelling hurricane growth. This image, taken by NASA's Aqua satellite, shows a three-day average for 25–27 August and highlights areas where temperatures reached around 28°C or above (red, orange and yellow), which is hot enough for hurricanes to form.

Katrina reached category 5 status before weakening and making landfall as a category 4 hurricane on 29 August. Meteorologists at Colorado State University in Fort Collins predict that there will be 20 named tropical storms this year. Katrina is the eleventh, and the hurricane season runs to the end of November.

The low-lying city of New Orleans escaped the annihilation that forecasters had feared from Katrina — Mississippi bore the brunt of the storm's force.

Hospital closure puts tissue bank in jeopardy

WASHINGTON DC

The world's largest tissue repository is without a home after a review panel decided to close the military hospital where it is based.

Scientists and pathologists met this week to try to influence the fate of the tissue bank at the Armed Forces Institute of Pathology (AFIP) in Washington DC.

The collection contains some 90 million samples, including those of rare tumours and diseases that can be used to develop new therapies, says David Seckinger, president of the Washington-based American Registry of Pathology. "The repository is unique," he says. "Its value to the American medical community is unparalleled."

The collection began in 1862 as part of a museum for disease specimens taken from Civil War soldiers. Over the years the institute expanded; it played key roles, for instance, in the study of the 1918 flu epidemic, Gulf War Syndrome, and in identifying victims of the terrorist attacks on 11 September 2001.

But the AFIP has long faced budget cutbacks (see *Nature* 424, 4–5; 2003), and earlier this year its host, the Walter Reed Army Medical Center, came into the crosshairs of the

Base Realignment and Closure Commission.

This independent commission periodically reviews operations at hundreds of US military bases and determines which should be closed. As part of the latest round of cutbacks, the commission voted on 25 August to permanently close Walter Reed's main campus in Washington. The panel recommended moving some AFIP functions, such as its forensics division, to Dover Air Force Base in Delaware, and preserving others, such as the tissue repository — but it offered no guidance on how to do the latter.

Without institute pathologists running the collection, it could fall into disuse and disrepair, says William Travis, a former AFIP pathologist now at the Memorial Sloan-Kettering Cancer Center in New York.

As *Nature* went to press, a coalition of pathology societies and other interest groups were due to meet to draw up a set of recommendations for the collection's future. The possibilities include attaching the repository to another government agency, such as the National Institutes of Health, or bringing it under the purview of a

public-private consortium. "There has been a flurry of different ideas and thoughts as to how it might be used," says Seckinger.

Some change might be good, says Renu Virmani, a former AFIP pathologist now at the International Registry of Pathology in Gaithersburg, Maryland. Currently, the collection is available for use only by the institute's 800 employees and their collaborators. A move could open the repository up for broader study.

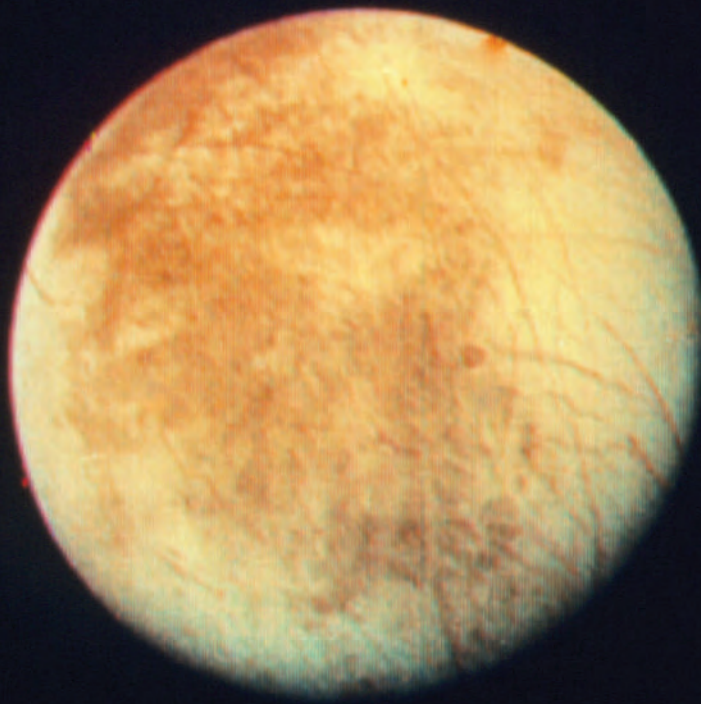
Ultimately, the Department of Defense will decide the tissue bank's fate. Travis, for one, is concerned that the department may try to warehouse the contents of the collection. "The repository is a living organism," he argues. "It is most meaningful in the hands of the top experts in the world."

Whatever happens to the AFIP, the Walter Reed centre is unlikely to be spared. The Base Realignment and Closure Commission will send its list of closures to President George W. Bush later this month. He is expected to sign it, giving Congress 45 days to reject the closures, which will otherwise become law.

Geoff Brumfield

"The repository is unique. Its value to the American medical community is unparalleled."

By Jupiter: the frosty satellite Europa features frequently in mission wish lists.



NASA/SPL

Designs on Europa unfurl

Mission designers at NASA may have found a way to explore Jupiter's moon Europa without busting the agency's budget — by flying past Earth first.

Europa has long been of interest to planetary scientists because of the ocean that is thought to lie beneath its icy crust, which may be a possible habitat for life. The National Academy of Sciences and other advisory groups have consistently listed Europa among the top destinations for future space missions.

But sending a spacecraft there is complicated by Europa's harsh radiation and the large amount of rocket fuel needed to brake into orbit. An earlier mission design from NASA's Jet Propulsion Laboratory (JPL) in California foundered in 2001 owing to cost and technical difficulty. And plans for a more ambitious nuclear-powered mission, the Jupiter Icy Moons Orbiter, have also been scrapped — at least for the foreseeable future.

But a study completed by JPL this summer has broken some of the previous barriers to visiting Europa. Work on the suspended nuclear mission led to progress in building radiation-resistant spacecraft components. And in setting ground rules for JPL's study, NASA eased a key restriction that was born of political concerns: mission designers were

allowed to send their plutonium-powered spacecraft past Earth and Venus to pick up propulsive energy before heading into the outer Solar System. NASA has come under fire from activist groups in the past for launching radioactive material into space.

The addition of Venus and Earth 'gravity assists' makes the trip to Jupiter longer, but allows a heavier spacecraft, with a substantial scientific payload, to launch on a single rocket. The Delta IV rocket in the study would still be an expensive ride — any Europa mission is expected to cost upwards of \$1 billion. But the mission may now fit within NASA's target budget for the first time.

The Europa Geophysical Explorer, as the concept is dubbed, could launch as early as 2012, carrying 150 kilograms of payload, including an ice-penetrating radar, a suite of remote sensing instruments and perhaps a small lander. The spacecraft would take more than six years to reach Jupiter and then spend a year-and-a-half orbiting the planet, including close fly-bys of Europa, Callisto and Ganymede, before ending with a 30-day intensive exploration of Europa. That would be long enough to map the subsurface ocean and

examine Europa's icy face at high resolution from orbit.

Scientists who have been lobbying for a Europa mission after the cancellation of the Jupiter Icy Moons Orbiter (see *Nature* 433, 342; 2005) hope the Europa Geophysical Explorer will make it into NASA's budget request as early as next year. This would allow work to begin in 2007.

That may be optimistic, given competing financial demands from the beleaguered space shuttle, the Moon-Mars astronaut programme and other science projects that have run into money troubles.

But the mission could get backing from NASA administrator Mike Griffin, who told a Senate committee in May that "You may look forward, in the next year or maybe even sooner, to a proposal for a Europa mission as part of our science line."

And international participation could help. Expectations are still high that any Europa mission will be done jointly with the European Space Agency (see *Nature* 434, 551; 2005). This is especially so after the successful cooperation on the Cassini-Huygens mission to Saturn.

Tony Reichhardt

"NASA has eased a key restriction that was born of political concerns."

Scientist quits climate-change panel

WASHINGTON DC

A prominent climate-change scientist has resigned from a US government panel, saying that colleagues tried to suppress his views.

Roger Pielke, of Colorado State University in Fort Collins, stepped down from the US Climate Change Science Program (CCSP) on 13 August. He says he quit in part because other panel members were trying to rewrite a report chapter he was charged with overseeing. "I was being prevented from including my views," he says.

Pielke is well known for his stance that other factors as well as carbon dioxide emissions cause climate change. The report is on temperature trends in the lower atmosphere, the first of 21 overviews commissioned by the CCSP. The group was set up by President George W. Bush to provide comprehensive reports on climate-change science for policy-makers.

Last month, some of the panel published papers in *Science* previewing parts of the report (see *Nature* 436, 896; 2005). The three papers suggest that observational problems are to blame for inconsistent measurements of atmospheric warming.

Pielke doesn't disagree with the published



Roger Pielke claims his views were edited out.

findings, but says that they narrow the remit of the report too much. "The CCSP charge was much broader than these three papers," he contends. The chapter was edited to bolster this narrowed view, he says.

Thomas Karl, who oversees the 22-member panel, admits that Pielke's chapter was edited. But he plays down the changes: "The new version was based on what had already been done — the changes were just an effort to push the process forward."

Pielke's departure is "unfortunate," says outgoing CCSP director James Mahoney, but his views will still be partly expressed in the final version. Pielke plans to submit detailed points during a public-comment period after the report is released later this autumn.

The resignation could have adverse effects, says John Holdren, an environmental-policy expert at Harvard University. "On a politically controversial issue like climate change, each little hiccup makes people wonder whether everything is in doubt," Holdren warns. "That perception can only be offset by a chorus of scientific voices saying that the findings are robust."

Geoff Brumfiel

NIH ethics rules come off probation

WASHINGTON DC

Six months after interim ethics guidelines shook up the US National Institutes of Health (NIH), permanent rules have been announced.

The regulations came into effect on 30 August. They ban NIH researchers from consulting for biotechnology or pharmaceutical companies, but work with trade or professional societies is allowed.

NIH chief Elias Zerhouni said last week that he would evaluate the ban on consulting in a year's time. In 2003, the *Los Angeles Times* unmasked researchers who accepted large consulting fees from companies whose products were being studied.

Under the new rules, no one at the NIH is allowed to hold shares directly related to their research. But most of them can now own unrelated shares in the biotechnology and pharmaceutical fields. Senior staff — "anyone with final decision-making

authority or next-to-final decision-making authority", according to Zerhouni — will have to limit holdings in each company to \$15,000, and to a total of \$50,000 for healthcare funds.

Ezekiel Emanuel, the chair of an advocacy group for NIH researchers, is cautiously pleased with the new rules. "I think they are certainly much more sensible than before," he says. But his group has no plans to disband: Emanuel says that many other policies still sap morale at the agency.

This week, Derek LeRoith resigned his job as chief of diabetes research at the NIH's National Institute of Diabetes and Digestive and Kidney Diseases, in part because of poor morale and because of the onerous interim ethics rules. "The rules are tough," he says. "They haven't loosened them as much as I thought they would."

Emma Marris


AMERICAN CHEMICAL SOCIETY BLOG

Catch up on news and views from the meeting in Washington DC

www.nature.com/news

Governors take the initiative over US carbon dioxide emissions

Nine US northeastern states have set targets for cutting back on regional greenhouse-gas emissions.

The move is a rare step forward for the United States, which in 2001 refused to ratify the international Kyoto Protocol for regulating emissions. It seems that in the face of federal inaction, individual states have begun making their own climate policy.

Under their plan, more than 600 power plants will cap their total carbon dioxide emissions at roughly 150 million tonnes — about what they emit now — starting in 2009. Between 2015 and 2020, they will cut back a further 10%.

In an arrangement not unlike the European Union's carbon-trading scheme, power companies will be given — or perhaps sold — the right to emit a certain amount of carbon dioxide. They may then sell credits if they cut emissions below that allocation.

"This is a big deal," says Judi Greenwald of the Pew Center on Global Climate Change in Washington DC, and an active participant in designing the policy. "It will be a really important policy experiment."

The plan comes from the two-year-old Regional Greenhouse Gas Initiative (RGGI), a coalition led by the governors of Connecticut, Delaware, Maine, Massachusetts, New Hampshire, New Jersey, New York, Rhode Island and Vermont. Together, the nine states emit as much carbon dioxide as Germany.

"Many national programmes start out with the states as labs," says Richard Valentinetti, director of air-quality control for Vermont and an RGGI participant. "Since the national programme was being discussed to death and nothing was happening, we felt we had to do something."

State leaders for air-quality and energy issues say they are about a month away from a memorandum of understanding, a handshake agreement on their goals. That, in turn, will beget a 'model rule' that each state would have to adopt in order to trade emissions allocations. Supporters hope that the rules could be law two years from now.

The northeastern pact is the furthest along of several US regional climate-change initiatives. Washington, Oregon and California have

Clear approach: individual states are planning a system that resembles the EU carbon-trading scheme.

similarly joined together in the West Coast Governors' Global Warming Initiative. California has adopted tough emissions regulations on its own: in June, Governor Arnold Schwarzenegger mandated a return to 1990 emissions levels by 2020, with a further reduction to 80% of 1990 levels by 2050.

In the northeast, the RGGI decided to take action in part because of the possible regional impacts of climate change on the states. "Most of them are coastal states — they are worried about rising seas," says Greenwald.

Temperatures are also an issue for states dependent on tourism. "Maine is worried about its forest. Vermont is worried about its sugar maples," Greenwald explains.

Details of the RGGI plan are still under debate, with environmental and industry groups wrangling — as might be predicted. One bone of contention is 'offsets'. These are separate actions, such as planting trees or capturing landfill gases, that companies could take in lieu of reducing emissions. The current draft suggests that offsets could make up half the difference between projected and capped

emissions, a number that some think is too high and others too low.

Cap-and-trade strategies have worked before in the United States, most notably in the 1990s when they helped lower sulphur dioxide emissions from power plants, reducing acid rain. But capping and trading carbon dioxide is not so simple, given the politically charged atmosphere of climate-change discussions.

A regional solution for a global problem makes no sense, says Bill Fang of the Edison Electric Institute, a leading industry group based in Washington DC. He also argues that capping the power industry alone is unfair. "Why should we be singled out?" he asks. The RGGI plan, for instance, makes no mention of transport, the second largest source of carbon dioxide in the United States.

James Brooks, director of air quality for the state of Maine, says that the effort is more of a message than a solution. "The RGGI region represents probably 3% of the world's greenhouse-gas emissions, so it is not going to have a huge impact," he says. "The idea is to set an example and to establish a working model that could be used nationally."

Emma Morris

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

R. PORTNER/STILL PICTURES

ON THE RECORD

“If evolution is true, why are there still monkeys?”

Popular US talk-show host Larry King interrogates philosopher Barbara Forrest on intelligent design.

“This deals a serious blow to the idea that the placebo effect is a purely psychological phenomenon.”

Jon-Kar Zubieta of the University of Michigan on his research that shows placebos can trigger the release of natural painkillers in the brain.

Sources: CNN, *Scientific American*

SCORECARD

**Iraqi marshes**

The United Nations finds that Iraq's devastated marshlands are making a “phenomenal” recovery. Being bogged down in Iraq is, in this case, a good thing.

**Hot beverages**

Forget fruit and vegetables — the number one source of antioxidants in the US diet is coffee. And if that's not to your taste, black tea came steaming in at number two, beating bananas and dry beans.

**Hairdressing**

The sprays, solvents and dyes used by hairdressers may be doing more than fashioning the latest styles for their clients. An occupational study reveals that members of the profession in the United States are among the most likely to develop Alzheimer's or motor-neuron diseases.

NUMBER CRUNCH

20 comets were discovered by Charles Messier, the eighteenth-century astronomer dubbed the “comet ferret” by King Louis XV of France.

32 comets were discovered by Carolyn Shoemaker, the world's most prolific comet-hunter using ground-based telescopes.

1,000 comets have been discovered by the Solar and Heliospheric Observatory spacecraft so far.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Satellite view alerts China to soaring pollution

Visitors to hazy Beijing can see how China's industrialization is fouling the air. Now data suggest that the situation is even worse than it looks, and pollution levels are rising.

Direct satellite measurements of a key pollutant — nitrogen dioxide — are reported in this issue (see page 129). The data show that concentrations of nitrogen dioxide in the atmosphere over China have risen by 50% during the past decade, and the build-up is accelerating.

In the 1990s, China introduced measures such as clean coal technologies to reduce air pollution (see *Nature* **435**, 1152; 2005). Estimates of nitrogen dioxide concentrations still rose by 13% between 1994 and 2000 — but there were hints of a plateau (D. G. Streets *et al.* *J. Geophys. Res.* **108**, 8809; 2003). The estimates were made as part of the ACE-Asia aerosol experiment and were based on ‘bottom-up’ calculations, which add up the fuel burned to gauge the pollutants released.

The satellite data come from the Global Ozone Monitoring Experiment (GOME), launched aboard a European Space Agency craft in 1995, and the Scanning Imaging Absorption Spectrometer for Atmospheric Chartography (SCIAMACHY), launched in 2002. Both experiments measure concentrations of trace gases in the atmosphere from the Earth's surface to about 10 kilometres high, although SCIAMACHY does so at a much higher resolution.

For nearly a decade GOME observations had shown increases in nitrogen dioxide over China at rates far greater than those estimated by the bottom-up measurements. But researchers didn't feel entirely confident about their results until they got the data from SCIAMACHY. “There was an element of pie-in-the-

sky to it,” says John Burrows at the University of Bremen in Germany, an author on the *Nature* paper. Burrows originally proposed both instruments: “We wondered, can you see this kind of thing? Now we know you can.”

China's emission inventories may have failed to take account of sources of pollution such as cars, whose numbers doubled in the country between 1995 and 2002. “New sources have stepped in to take the place of old ones,” says another member of the Bremen team, environmental physicist Andreas Richter.

The team's calculations depend on an assumption about how nitrogen dioxide concentrations vary vertically in the atmosphere. “But this should not affect measurements of trends,” Richter says.

“The satellite observations are a good starting point to tell us where to correct the predictions,” says Tami Bond, an environmental scientist at the University of Illinois, Urbana-Champaign, who worked on the ACE-Asia inventory. However, she adds, “they don't tell you exactly what is happening”.

The extent of the increase surprises Jianzhong Ma, an atmospheric chemist at the Chinese Academy of Meteorological Sciences in Beijing, who next year will begin a study of air pollution and aerosols. He hopes that ground-level observations and aircraft sampling will pin down exact amounts of nitrogen oxides. “We need to integrate the methods,” he says.

Meanwhile, the satellite researchers say they will hone their data. GOME II is scheduled for launch next autumn, and Burrows is already proposing a geostationary satellite that could observe continuously, reducing uncertainties about daily fluctuations.

David Cyranoski



**SURGING HORMONES
BLAMED FOR PAIN**
Study of sex-change
patients reveals role
of oestrogen.
www.nature.com/news

Commuters remain stoical in face of terrorist attacks

Londoners are coping relatively well after the terrorist attacks of 7 July, psychologists say.

In a survey of 1,010 Londoners, 31% reported symptoms of substantial stress — a number that is somewhat lower than has been found in other countries following terrorist attacks. The study was conducted after the 7 July bombings but before the failed attacks of 21 July.

Muslims, who may have felt singled out in the search for perpetrators, grouped with those who had struggled to contact family or friends immediately after the bombings as the people who experienced the most stress.

The findings appeared

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Very few of those caught in the London bombings want counselling.

online last week in the *BMJ* (G. J. Rubin *et al.* doi:10.1136/bmj.38583.728484.3A; 2005).

After the 11 September 2001 terrorist attacks in the United States, 44% of the people surveyed in New York and Washington reported feeling substantial stress.

Although the London attack was on a much smaller scale, Londoners may have fared better because they have a history of dealing with terrorism, says Neil Greenberg, a trauma expert at King's College London and co-author of the recent study.

In New York, 11% of those questioned went on to develop post-traumatic stress disorder (PTSD) within one to two

months. Psychologists have found even higher rates of PTSD in the Middle East (as much as 40% in one study) where repeated violence creates an ongoing sense of threat.

Greenberg says it is too early to tell if Londoners are suffering from PTSD, but his team's results suggest that this is unlikely. "If one third were substantially stressed less than two weeks after the bombings, then it is reasonable to think that the level of PTSD will be less than this," he says. "We know that people's distress levels fall rapidly with time."

The team plans to interview the same people in six months to see how they are doing. ■
Jennifer Wild

Journal urges honesty about homeopathy's flop

Homeopathy offers no greater benefit than the placebo effect, according to an extensive review in *The Lancet*.

The study, published on 26 August, compares 110 placebo-controlled, randomized trials of homeopathy with 110 trials of conventional medicine for the same disorders. Matthias Egger of the University of Bern, Switzerland, and colleagues report that although smaller studies revealed benefits from both types of treatment,

larger trials revealed that only conventional medicines had a convincing impact (A. Shang *et al. Lancet* 366, 726–732; 2005). The scientists suggest that any positive effects of homeopathic treatment may stem from the close bond that often develops between practitioner and patient.

In its editorial, *The Lancet* says, "Doctors need to be bold and honest with their patients about homeopathy's lack of benefit, and with themselves about the failings of modern medicine to address patients' needs for personalised care."

NIH aims for online-only grant applications

The US National Institutes of Health (NIH) is to make its grant application process entirely electronic, the agency announced on 19 August.

Applications will be made through www.grants.gov, a website that serves as an access point for 26 federal grant-making agencies. By October 2006, all grants that involve a single research project will be electronic. Other grants will follow by the end of 2007. The NIH hopes that paperless applications will shorten the time between grant application and decision, which is

currently about 9 months. The National Science Foundation's existing online application system, FastLane, has been praised by researchers for helping to reduce problems with formatting applications.

Japan speeds ahead with plans for supersonic plane

An unmanned supersonic plane will be given a test flight in the skies over Woomera, Australia, in the next few weeks. A previous flight three years ago failed when the plane separated from its rocket booster shortly after launch and crashed.

Engineers at the Japan Aerospace Exploration Agency plan to launch an improved version with the aid of a rocket booster. The plane will leave the rocket at an altitude of more than 20 kilometres and fly at twice the speed of sound for 15 minutes. The agency wants to verify the computational fluid-dynamic techniques used to develop the craft's wings.

Success of the ¥10 billion (US\$90 million) project, which began in 1997, is important for Japan as it aims to earn its place in the international collaborations expected to form as nations seek to develop next-generation supersonic transport.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Drug problem: a large-scale survey finds that homeopathic medicines have no real benefit.

Libyan and US nuclear weapons labs join forces

Researchers from US nuclear weapons labs will soon collaborate with their Libyan counterparts. Under a new agreement announced on 24 August, researchers from three US labs, including the Los Alamos National Laboratory in New Mexico, will work with scientists from the former Tajura Nuclear Research Centre near Tripoli.

The move is the latest in a series of research agreements between the US and Libyan governments following the lifting in 2004 of sanctions against the North African state.

Although not specifically intended to target researchers associated with Libya's former nuclear weapons programme, it won't exclude them either, says Bryan Wilkes, a spokesperson with the National Nuclear Security Administration, the US agency overseeing the cooperative agreement.

World Bank seeks to turn the tide on overfishing

Unsustainable fishing practices are to be targeted by a team of policy analysts led by the World Bank.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Net benefit: preventing overfishing will preserve stocks for the next generation of fishermen.

The Profish programme, which was launched on 24 August at the Fish for All Summit in Abuja, Nigeria, will involve a group of experts visiting countries on request and working with governments to tackle measures such as subsidies that encourage large fishing fleets. The team will also help create and maintain the 'opaque list', a register designed to expose vessels whose activities are not transparent and clearly above board.

The money for the project comes from the World Bank, Norway, Iceland, France and Finland, and amounts to just US\$1 million so far, although the World Bank says more will soon be raised.

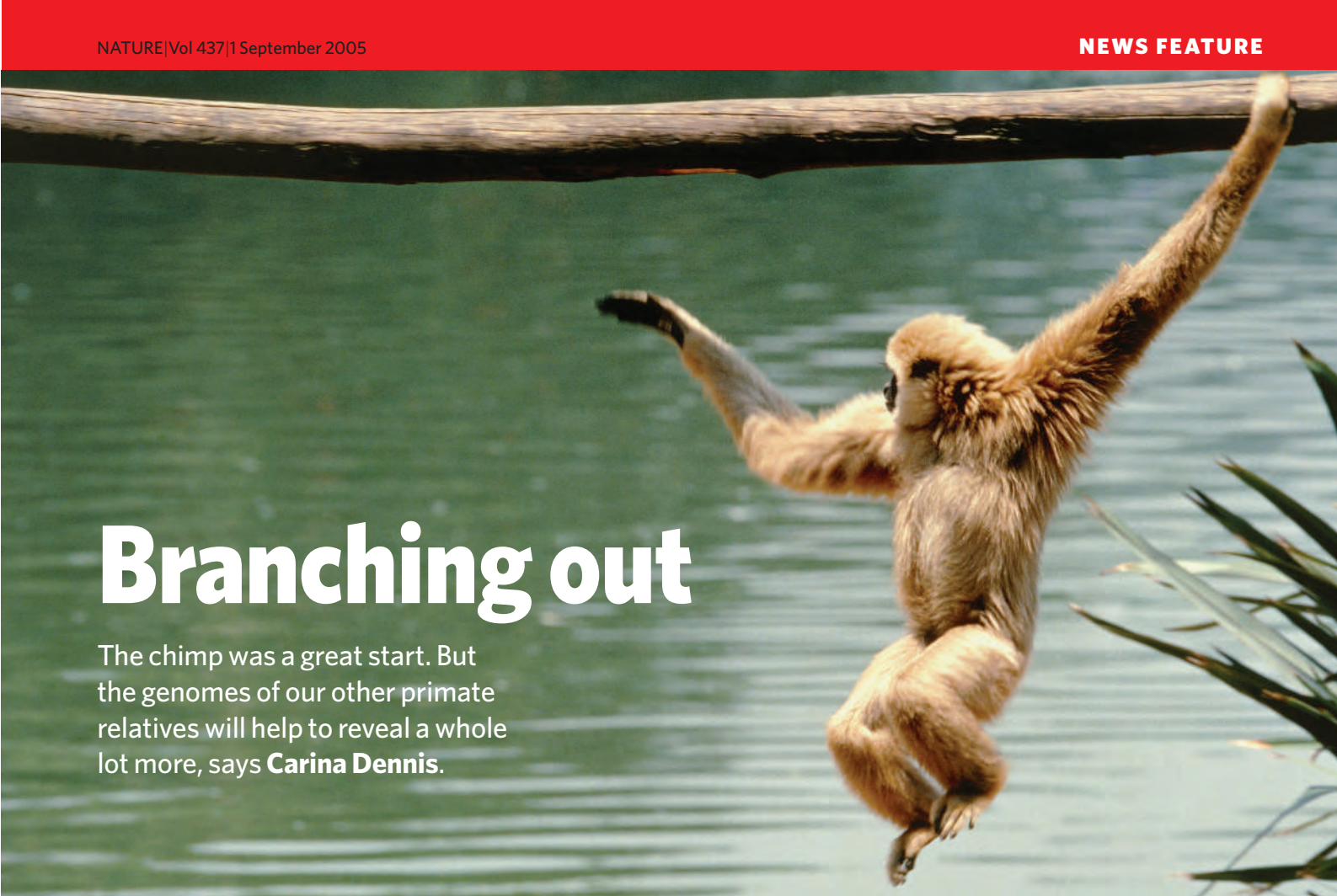
Hamster study offers hope of blood test for prions

A blood test for variant Creutzfeldt–Jakob disease, the human form of mad cow disease, may soon be a realistic possibility.

The disease is thought to be caused by misshapen prion proteins that clump together and damage sufferers' brains. A test for prions has proved tough to develop, as the proteins accumulate in the brain and are only present in small quantities in blood.

A team led by Claudio Soto of the University of Texas Medical Branch in Galveston has now created a technique for amplifying prions in hamster blood. The researchers mix normal prions with infectious ones and use sound pulses to break up the clumps that form, allowing the free prions to repeatedly infect the normal versions (J. Castilla *et al. Nature Med.* doi:10.1038/nm1286; 2005). Soto says it will take about six months to establish whether the technique works in humans.

G. OSODJI



Branching out

The chimp was a great start. But the genomes of our other primate relatives will help to reveal a whole lot more, says **Carina Dennis**.

“It’s a dream come true,” says Carole Beth Stewart of the chimp genome. The evolutionary biologist from the University at Albany in New York State never imagined that primate genomes would be sequenced in her lifetime. She can barely contain her excitement, not only over the draft chimp genome sequence¹ but also about those of the other primates, including the orang-utan and rhesus macaque, which will soon be available.

Stewart is one of many who hope that looking at several primate genomes will help answer fundamental questions about our own evolution and that of other primates (see graphic, overleaf). What underlies the differences between humans, apes and the other primates? How did the physical structure and content of our ancestors’ genomes enable primates to evolve the way they have? Will this give us more insight into how evolution itself operates?

Chimps, our closest living relatives, are a great starting point. But our genomes are too much alike to get meaningful answers to many of these questions. “It’s frustrating that humans and chimps are so similar,” says Andrew Clark of Cornell University in Ithaca, New York. It’s difficult to tell whether a DNA sequence in humans that is missing in chimps was really added during human evolution or has simply been lost in the chimp lineage.

Another problem is that it is hard to be sure

straight away that any differences found are significant. “You find a difference that you think could be very exciting, but it could just turn out to be a natural variant within one species,” says Ajit Varki at the University of California, San Diego. Chimps, like humans, differ genetically from each other, although the extent is debatable. More chimps from different subspecies must be sequenced to capture the full extent of sequence diversity. And the chimp genome sequence is still only a draft. To ensure that the differences found are real, the chimp sequence needs to be improved to match the polished ‘finished’ standard of the human genome. This is now under way.

Even so, researchers need other primate genomes if they are to address the question of which genetic changes are unique to humans or chimps (see page 50). The rhesus macaque, an Old World monkey, will be the first available — a preliminary assembly of its genome sequence was released into the public databases earlier this year and an improved version is expected by the end of the year. The push to sequence its genome stems from its popularity in biomedical research². It will help researchers figure out whether the differences arose in the lineages leading to modern chimps or humans after they split from their last common ancestor approximately 6 million years ago.

But although the macaque is a useful reference, it is not ideal for identifying genetic changes that happened after the human-

chimp split, as it diverged from a common ancestor some 25 million years ago. “There have been so many changes, it will be harder to tell what’s gone on,” says Varki.

To better understand how the human genome has evolved, researchers want to look at a primate that is sufficiently different from humans and chimps, but which shares a more recent common ancestor. “The obvious one is the orang-utan,” says Varki. The orang-utan, a great ape like the chimpanzee, diverged from a common ancestor with chimps and humans approximately 12 million years ago. Its genome is currently being sequenced and a draft is expected early next year. “Identifying sequences common to human, chimp and orang-utan, but different in the rhesus monkey, would provide valuable clues to the genomic features distinguishing great apes from other primates,” says Eddy Rubin, director of the Joint Genome Institute in Walnut Creek, California.

But for others, the most exciting primate genome to follow chimp will be that of the gorilla. This is our next closest primate relative, and some parts of the gorilla genome are closer to humans than is the chimp genome³. “The sequence will help us understand how the species formed that went on to become gorilla, chimp and humans,” says Svante Pääbo of the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany.

Generally speaking, the closer a primate

T. MCHUGH/SPF

sequence is to human, the more useful it is for figuring out more recent, human-specific traits. And the more species that can be compared the better. If orang-utan, gorilla and chimp were all identical at one DNA position and humans were different, for example, then geneticists could be quite confident that that change happened in the most recent history of the human lineage. "We expect to start sequencing the gorilla in October this year," says Jane Rogers of the Wellcome Trust Sanger Institute near Cambridge, UK. A draft assembly should be available in a couple of years, she adds.

Back to our roots

While some researchers are working on the the youngest shoots of the primate family tree, others are delving at the roots, to understand what the earliest primate genomes were like. To this end molecular palaeontologists are keen to sequence representatives from each of the major primate lineages. The sequencing of the marmoset, a New World monkey, has just begun. "I would also like the lemur sequence," says Asao Fujiyama of the National Institute of Informatics in Tokyo, Japan, who was part of the team that sequenced the first chimpanzee chromosome last year⁴. The suborder of primates to which lemurs belong arose earlier than the branch leading to monkeys and apes. "My interest is to trace how modern human chromosomes have evolved from our ancestor," he says.

David Haussler of the University of California, Santa Cruz, and his collaborators want to peer even further back in time⁵. They are analysing sequence data from across the animal kingdom to reconstruct the genome of the ancestor of placental mammals, which lived around the time of the dinosaurs more than 75 million years ago. "Our goal is to reconstruct the complete history of the DNA changes from the placental ancestor to the modern human," says Haussler. Reconstructing this genome and comparing it with the human sequence will make the key genetic changes in our evolution from that ancestor much easier to see, he explains. "Additional primate genomes will help fill in the missing details." His team is now assembling the first draft of the ancestor's reconstructed DNA sequence and, although preliminary, the results show that it is at least computationally feasible, says Haussler.

Clambering back up the tree will add to the picture of how genomes evolve and how the genes within them work. The gibbon, which shares a common ancestor with the great apes, has a most peculiar genome, according to Todd Disotell, a molecular anthropologist at New York University. Its chromosomes seem to have changed and evolved faster than those of other apes. "Its genome looks like it has been put in a blender," says Disotell. It seems to have virtually the same DNA content as humans and chimps, but all churned up. It will be inter-



esting to find out whether the functions of genes change in their new chromosomal locations, says Disotell.

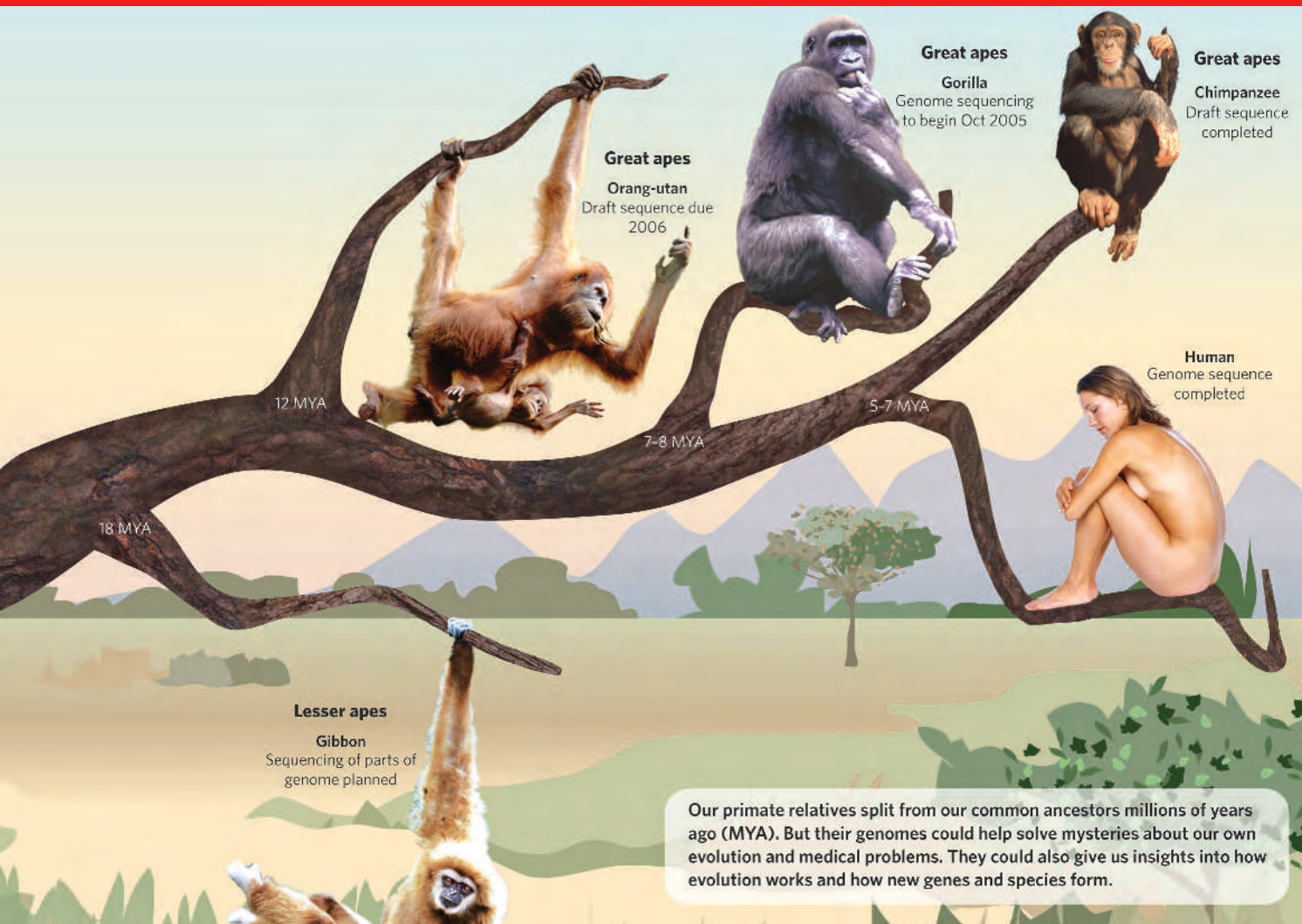
William Murphy of Texas A&M University is also excited by the genomic clues thrown out by gibbons. His team has reconstructed the chromosomal architecture of a mammalian ancestral genome by comparing stretches of genomic sequence from eight very different mammals. The results suggest that stretches of duplicated sequence promote chromosomal rearrangements. In turn, these contribute to genetic changes that can lead to new species. If the same holds true in the gibbon, we might get a better handle on the mechanism of genome rearrangement, says Murphy.

When a chromosome breaks and rejoins, clues to the mechanism and molecular machinery involved can be left behind in the sequence. Because the gibbon genome contains so many rearrangements, it might be easier to identify the tell-tale footprints of the machinery involved. Earlier this year, the US National Human Genome Research Institute in Bethesda, Maryland, announced it would

fund the sequencing of small portions of the gibbon genome to capture some of these rearrangement sites.

Structural changes like these may have been important in driving human evolution. The draft chimp genome revealed that 2.7% of its genome differs from humans because of duplications, compared with 1.2% differing at single base-pairs⁶. Structural changes such as duplications are often hotspots for the birth of new genes. But right now it's impossible to tell whether single base changes or structural variations have had the biggest influences on how we evolved. "I couldn't hazard a guess," says Evan Eichler of the University of Washington in Seattle.

The emerging picture is of primate genomes that have been shaped in a variety of ways. Some changes were single-base alterations; others were 'structural variations', such as insertions, deletions or duplications of sequence. And, periodically, a transposable element — a parasitic DNA sequence — would infect and spread through the genome, tending to pool in the non-coding regions. In



PICTUREARTS/NEWS.COM; CREDITS: A. ROUSE/NHPA

addition, a whole genome doesn't change at an even pace; comparisons of many primate sequences will reveal how different genomic regions have evolved at different rates.

Beyond evolution

Primate genomes can give us much more than a fascinating history lesson. They are, for example, providing valuable insights into human disease. Rubin has devised a method for comparing similar genomes and picking out functional genes and control sequences from 'junk' DNA. Dubbed 'phylogenetic shadowing', the technique has let him compare numerous different primate DNA sequences (including human), and to spot stretches of DNA that have remained broadly the same throughout relatively recent evolution. This suggests that the correct sequence of these regions is so important for the survival of the animal that evolution cannot tinker with it. Rubin's team first used this approach to discover primate-specific stretches of sequence that control the production of the protein apolipoprotein A, whose faulty regulation is implicated in susceptibility to atherosclerosis⁷. They are now looking for important regulators

of the gene for the low-density lipoprotein receptor, which is involved in controlling blood cholesterol.

The good news, says Rubin, is that only a handful of carefully chosen primate genomes are needed to identify the most interesting genetic elements. The phylogenetic spread that would capture most of the genetic diversity in primates, he adds, would be — in addition to human — the Old World monkeys rhesus macaque and colobus and the New World monkeys marmoset, titi and spider monkey.

Molecular biologists aren't the only ones who hope to benefit from the chimp genome. The ancestors of most primates — unlike those of humans — seem to have left behind few fossils (see page 105), probably because they died in environments unfavourable to fossilization. "We are very fortunate that humans had the decency to evolve in good places for preserving fossils," says David Penny of Massey University in Palmerston North, New Zealand. This means that there are lingering questions over when certain primate lineages diverged, the size of populations at the time of the splits, and phylogenetic relationships among the more than 60 genera of living primates⁸. More primate genome sequences will help to calibrate the times of divergence and resolve phylogenetic discrepancies.

And genome sequences have limitations. "You can only learn so much from the genome sequence," says Penny. To make sense of the sequence differences between primates, researchers need information on the expression of genes in different tissues and the genetic variation in family pedigrees and different populations. Obtaining samples from these endangered animals in an ethical way is hard, say researchers (see page 27). "There are bits of dead gorillas in freezers but they're not great to use," says Rogers. With wild apes threatened with extinction, it is imperative to collect blood and tissue from captive populations and from animals that die in the wild. "The opportunity is fast disappearing," says Eichler. "We have only a short window to act in."

Carina Dennis is Nature's Australasian correspondent.

1. The Chimpanzee Sequencing and Analysis Consortium *Nature* **437**, 69–87 (2005).
2. Cyranoski, D. *Nature* **418**, 910–912 (2002).
3. Chen, F.-C. & Li, W.-H. *Am. J. Hum. Genet.* **68**, 444–456 (2001).
4. The International Chimpanzee Chromosome 22 Consortium *Nature* **429**, 382–388 (2004).
5. Blanchette, M., Green, E. D., Miller, W. & Haussler, D. *Genome Res.* **14**, 2412–2423 (2004).
6. Cheng, Z. *et al. Nature* **437**, 88–93 (2005).
7. Boffelli, D. *et al. Science* **299**, 1391–1394 (2003).
8. Goodman, M., Grossman, L. I. & Wildman, D. E. *Trends Genet.* **21**, 511–517 (2005).



W. WALLAUER/JG(LEFT), J. MOLLISON (RIGHT)

Emotion picture: James Mollison's image of Chim tries to remind us that apes are close kin.

What the chimp means to me

Interacting with our closest living relative can be a profound experience. To mark the publication of the chimpanzee genome, *Nature* asked four individuals for their different perspectives.

James Mollison: Picture this

The chimp genome reinforces just how close we are to our primate relatives. "But I have never doubted the similarities between human and chimp," says James Mollison. Staring into the eyes of a chimpanzee's face, its photo blown up to an impressive two metres tall, you can see where he's coming from.

Mollison's photographic exhibition of ape portraits, called *Face to Face*, is currently showing at the Natural History Museum in London. His pictures of 30 chimpanzees, bonobos, gorillas and orang-utans are all taken in passport style — to reinforce a feeling of kinship when people view them. "That's the one photo that everybody has," Mollison says.

This required the photographer to get close to his subjects, which put wild animals out of



the picture. So Mollison started phoning and e-mailing wildlife sanctuaries. Those in his native Britain were not very supportive. "They said you can't expect a chimp to sit for a portrait," he says. But over the course of three years, Mollison completed his project by visiting zoos and ape sanctuaries in Africa, Asia, Europe and the United States.

You can get a chimp to pose, Mollison found — although it's difficult. First he had to gain their trust, spending days hanging out with them and making friends. "Getting them to stare into the camera wasn't easy," he says. One trick was to pick an imaginary flea from their chin, then put it on top of the lens. Bribery with peanuts also worked well.

From the crow's feet creases in the apes' faces to their grey hairs and double chins, the

resulting images appear uncannily human, each face as individual and expressive as our own. Staring into their chocolate-brown eyes, the full gamut of human emotion seems to stare back. The wrinkled faces portray, by turns, humour, anger and wisdom.

Mollison is known for his photojournalism. He has depicted the ravages of tuberculosis for the World Health Organization, and when working for Benetton's *Colors* magazine his subjects included La Modelo prison in Bogotá, Colombia, and East African refugees. But he admits to having known little about the suffering wrought by the bushmeat trade until he embarked on the *Face to Face* project. Most of Mollison's subjects were orphaned when their parents were slaughtered for meat. And years later, many still suffer from emotional trauma.

Chim, for instance, was photographed in 2001 at the Mvog Betsi Zoo in Cameroon. As an infant, her parents were killed by poachers. Later 'rescued' by a local environmental journalist who dressed her as a child, Chim was taught to dance to receive food — something she still does, when hungry.

Despite their troubled pasts, most of the animals have retained a sense of mischief, says Mollison. Some apes dive-bombed him from trees; others untied his shoelaces. "They tried to break anything they could. The level of excitement reminded me of being in a pillow fight when I was a kid. But they were also incredibly warm, and needed affection."

Mollison hopes his work will further the cause of ape conservation, and sees it as a celebration of evolution. The faces and expressions portrayed in his pictures are so similar to ours that they blur the boundary between human and ape. With about a quarter of US college graduates denying that humans and chimps have a common ancestry, this is an important message to convey, Mollison says.

Helen Pilcher

♦ www.nhm.ac.uk/face-to-face

Peter Singer: Rights and wrongs

D. APPLEWHITE/PRINCETON UNIV.

"We demand the extension of the community of equals to include all great apes: human beings, chimpanzees, bonobos, gorillas and orangutans." So reads the *Declaration on Great Apes*, a statement issued by the Great Ape Project.

Co-founded in 1993 by the Australian-born ethicist and philosopher Peter Singer, the project's ultimate goal is for chimps and other great apes to be granted three 'human' rights: the right to life, to liberty and to protection from torture. "The fact that they clearly have some self-awareness shows that we should treat them differently," says Singer. "The case for granting them some basic rights is a stronger one than might be made for mice and other animals."

Singer believes the Seattle-based project has influenced reforms enacted over the past decade. Chimpanzees are used far less often in invasive biomedical research than they used to be, and when they are too old or sick to be used in research, scientists now retire them to sanctuaries, instead of killing them. "I think we've had some impact in spreading this consensus," he says.

Singer is a controversial figure, whose views on animal rights, abortion and euthanasia have won both plaudits and violent criticism. An advocate of veganism and opponent of



most vivisection, his 1975 book *Animal Liberation* is widely credited with launching the animal-rights movement. Now based at Princeton University in New Jersey, he has outraged some religious groups with his support for abortion, and his justification of euthanasia in cases where a patient, such as someone overtaken by Alzheimer's disease, has

become a "nonperson".

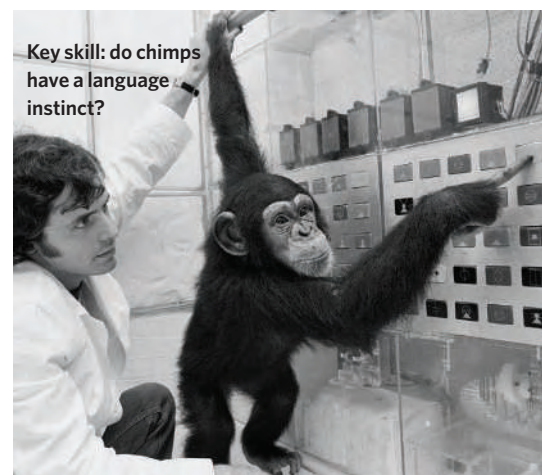
At the heart of Singer's ideas lies a utilitarian approach to ethics, and a rejection of 'speciesism'. The striking genetic similarity between people and chimps is not a crucial factor in shaping this outlook, he says: "I don't think that knowing which genes chimps share with us actually determines anything about their moral status in any meaningful way."

Singer believes that apes' rights come from their moral and cognitive capacities. But he hopes that the publicity surrounding the publication of the chimp genome will advance the Great Ape Project's cause, and drive a greater respect for all animals — not just chimps.

"It will help bridge the gulf that we mentally place between ourselves and animals," he says. "We will see chimpanzees as kin, and the differences between us and other animals as graduated, rather than a sharp discontinuity." ■

Erika Check

♦ www.greatapeproject.org



Key skill: do chimps have a language instinct?

Gary Marcus: Talking point

BETTMAN/CORBIS

Give a chimpanzee a keyboard, a computer and years of tuition, and the betting is it won't write the complete works of Shakespeare. In fact, you'd be lucky if it stumbled on a sentence. Chimpanzees may be many things, but linguists they are not, says Gary Marcus, a cognitive scientist at New York University.

Scientists and the public alike have been impressed by the talents of animals such as the chimp Washoe and the bonobo Kanzi, who have learned to communicate using sign language and keyboards. But Marcus remains sceptical of the value of such experiments. "It's a silly game to see how much a chimp can act like a human," he says.

No one can doubt chimpanzees' ability to communicate. In the wild, the animals grunt, screech and holler — vocalizations that are part of the fabric of chimp society. And in the lab, chimps can learn and use symbols in a way similar to that in which young children use words. But is this really language?

Marcus thinks not. Chimps learn words one at a time and never show the explosive acquisition of language accomplished by excited toddlers. And although children learn to talk about past, present and future, chimps seem to communicate solely about the here and now.

Chimps also lack what some experts have called the linguistic 'silver bullet': the ability to combine bits of language into larger units. Recursion, as it is known, expands the range of possible topics and lets the speaker appreciate the views of others. Even the most sophisticated chimp, Marcus points out, would be baffled by a sentence such as: "She knows that I know where the peanut is hidden."

Humans may be the only animals to crack



AP PHOTO/P. SAKUMA



Ethicist Peter Singer believes that apes should be granted the right to life and liberty.

the recursion nut, but the skill is just one of many likely to be crucial for language, says Marcus. And the chimpanzee genome may help us to pinpoint others.

"Evolution tends to proceed not by starting over but by tinkering with what is already in place," says Marcus. Although chimps don't have language, they're likely to share some of the features that predispose one to it. Knowledge of the genetic sequence may help us decipher the genes behind this shared cognitive backbone.

"In the meantime," says Marcus, "it will help if people stop worrying specifically about whether chimps happen to have language and instead ask: what are the many things we have in common, and how did those pave the road to language?"

Still, Marcus understands why some people want to ascribe human-like abilities to chimpanzees. "I think it's pretty hard not to anthropomorphize chimps," he says. "Our brains are set up to analyse other entities in terms of

their goals, beliefs, desires and so forth, and chimps look for all the world like they've got those things."

He does not, however, think that those working with chimps should become completely detached and dispassionate: "Good scientists are objective, but objectivity does not demand that scientists be blind to potential points of contact between humans and other species."

Helen Pilcher



Although the bushmeat trade is a threat to chimp numbers, the Ebola virus is also taking its toll.

would take 150 years to recover, he says.

Vaccines are an option. Two candidates that protect lab monkeys — and may protect apes — already exist, but these are being developed for human use. Last year, the US government set aside US\$6 billion to speed the development of drugs and vaccines against bioterror agents such as Ebola. But funding for ape studies is harder to come by. Just US\$10 million would fund small-scale lab tests and pilot studies on wild apes, says Walsh, adding that the vaccine could make it to the field within a few years. "In terms of conservation, it's a bargain," he says. But it would require a concerted lobbying campaign that — to Walsh's immense frustration — shows no sign of emerging.

In the meantime, simple measures could make a difference. Ebola is thought to be spread by an animal — as yet unidentified — which doesn't succumb to the severe disease. Perhaps this creature doesn't like wet feet, as outbreaks appear to be confined by water. Clearing small rivers of overhanging trees might halt the disease's spread, says Walsh. "All you need is a dugout canoe and a chainsaw."

Law enforcement also needs to be stepped up to protect apes from hunters, says Walsh. Realistically, just a handful of west Africa's score of national parks have effective anti-poaching strategies. Elsewhere, organized gangs of hunters supply a well-structured food chain transporting bushmeat to major cities. "The poachers get cigarettes and a few hundred francs," says Walsh. "It's their bosses who are making the profit." Some of the ringleaders do get arrested, but corruption is rife, so they often walk free.

Unless things change, ape populations will continue to dwindle. Currently, there are about 15 sites that host 5,000 or more great apes. Over the next ten years, Walsh predicts these could shrink to just a couple of strongholds, each with a couple of hundred apes or fewer.

The chimpanzee genome may yield many things, including therapies to treat ailing apes, "but it will have zero practical impact on chimpanzee conservation", warns Walsh. "The repercussions of the chimpanzee genome will arrive in Africa in 10 to 15 years time. By then it will be too late."

Helen Pilcher

Peter Walsh: Going ape

Like many of today's conservation biologists, Peter Walsh was drawn to his chosen career by television documentaries showing Jane Goodall and the chimpanzees of Gombe in Tanzania.

Today, those cosy childhood memories are overshadowed by a sense of desperation and outrage at the plight of Africa's remaining chimps. Frustrated by politics and plans that don't deliver, Walsh is also battling against sceptics who doubt his evidence of a looming catastrophe. "The world's great apes are in serious danger. If we don't act fast, it's going to be too late," he says.

Walsh, who works at the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany, has reason to be alarmed. His modelling studies suggest that ape populations in western equatorial Africa — the world's last stronghold for chimpanzees and gorillas — have been slashed by more than half in the past 20 years (P. D. Walsh *et al. Nature* **422**, 611–614; 2003).

Walsh blames illegal hunting and a serious epidemic of the deadly Ebola virus. Although



biologists acknowledge the bloody impact of the bushmeat trade, he is frustrated that few seem worried by the insidious threat from Ebola.

"I'm totally amazed by the response," he says. Some experts even deny that ape populations are in crisis at all. They presume that because western equatorial Africa's forests are still intact, the resident ape populations are safe. Walsh says

they don't bargain on the ravages of natural disease and the dogged determination of poachers.

Many conservationists think that ape populations will simply bounce back from the virus's attack, as resistant animals interbreed and repopulate the forests. But with female chimps taking 14 years to reach sexual maturity and producing just one baby every six years, that's not going to happen, warns Walsh. Left to their own devices, some populations

"The world's great apes are in serious danger. If we don't act fast, it's going to be too late."

BUSINESS

US set to endorse human pesticide testing

Pesticide manufacturers are pushing hard to ensure that data from toxicity tests on people can be used in licence applications for their products. **Meredith Wadman reports.**

The pesticide industry is keenly awaiting an imminent US ruling. Within days, the Environmental Protection Agency (EPA) is expected to publish its proposed policy on the use of human subjects in tests to assess the safety of pesticides.

The \$10-billion industry has plenty riding on the details of this regulation. It wants to use human toxicity data to help it keep at least a dozen pesticides on the market. These products must pass new safety hurdles with the EPA by next August or face being banned or restricted.

Perhaps as important, a formal EPA policy governing company-sponsored human tests would give an unprecedented government seal of approval to the controversial practice. Critics predict that this will encourage companies to conduct human toxicity tests on a wide range of some 1,200 active pesticide ingredients now on the market, with the aim of loosening their regulation.

Pesticide manufacturers say that human data are vital for fair, science-based regulation of their products. Tests on humans "provide very valuable insights into exactly what happens when humans are exposed to low levels of a compound", argues Ray McAllister, an agronomist and policy analyst at CropLife America, a Washington-based lobby group that represents the US pesticide industry. This, he says, reduces the uncertainty inherent in relying only on animal tests.

Independent scientists have backed the industry, with careful caveats. Most notably, the National Research Council in a 2004 report concluded that human test data could be used by the EPA, if strict ethical and scientific standards were met (see *Nature* 427, 770; 2004).

But environmental groups such as the Natural Resources Defense Council (NRDC) are scathing in their criticism of human experiments. They say that the industry wants to conduct them not for the public good, but to keep threatened products on the market. They point out that the use of human data could allow the industry to neutralize additional requirements dating from 1996 that govern the allowable levels of pesticide residues on food. Those rules

set the levels ten times lower than their previous value — the human data could reverse that change for many products, experts say.

Some scientists also argue that there are few circumstances in which such studies can be ethical, and that the EPA should not accept data from them.

"Human dosing studies have failed to meet widely accepted ethical standards for the conduct of research," says Alan Lockwood, a neurologist at the University at Buffalo, New York, who has closely examined six of the company-conducted studies. "There is no assurance that any such study can be completely free of risk."

The human factor

The new EPA rule — which will remain provisional during a 90-day period of public comment — will mark a key moment in a decade-long political battle between pesticide makers and their critics. Congress joined the fray this year, when two Democrat senators blocked the nomination of Stephen Johnson as EPA administrator until he agreed to cancel a study on children.

Last month, a law came into effect that commands the EPA to finalize the rule within six months, and forbids it from considering human data in the interim. That bill was seen as something of a victory for pesticide makers, as they managed to head off a one-year moratorium sought by critics.

The current debate stems from a 1996 law that made it much harder for companies to meet EPA safety standards using animal tests alone. The Food Quality Protection Act was intended to tighten standards to protect vulnerable people such as children and babies from pesticide residues on food. But it encouraged pesticide makers to resume some human tests — which they had largely

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Pesticide makers say that many of their products would be more fairly regulated if human toxicity data were taken into account.

abandoned — and to resurrect data from studies dating as far back as the 1960s.

Concerned by this turn of events, the EPA launched a de facto moratorium on the use of human test data in 1998, which was formalized in 2001. The pesticide makers sued the agency and won, and since 2003, the EPA has considered human data on a case-by-case basis.

As of June this year, 24 human studies had been presented to the EPA by pesticide companies. Most have not been published, but opponents have criticized their ethics after the EPA produced them in response to demands from Congress. Some bioethicists have also criticized what they see as deficient informed-consent procedures, inadequate statistical power and financial conflicts of interest.

Several of the studies, for example, were done in Britain by Bayer, the world's largest pesticide manufacturer. In 1998 and 1999, it paid contractors to conduct three human dosing experiments using its pesticide azinphos methyl, an organophosphate that is used on 73% of US apples and more than half of US pears, cherries and blueberries. In high doses, it can

lead to convulsions and death.

In a letter to one potential study participant,

D. WILSON/CORBIS

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

J.W. ROGERS/CORBIS

the pesticide was described as a “drug” that would be “administered orally”. In a volunteer information sheet for another of the studies, nausea, vomiting and stomach cramps were listed as possible side effects, but weakness, respiratory failure and death were not.

Bayer declined to comment on the particulars of its studies, but in the past it has defended them as meeting international ethics requirements.

The studies have formed an important part of Bayer's pitch to save azinphos methyl from the axe under the tighter safety rules of 1996. In 2001, the EPA banned dozens of uses of the pesticide, but allowed its application to 15 crops to continue. It is those applications that the company is now fighting to retain.

Last November, in a key battle in this struggle, a dozen Bayer officials met with 15 EPA regulators at the company's request. Bayer's officials used the British studies to argue that the allowable levels for people's exposure to the pesticides' residues should be 17 times higher than they are at present. The EPA's decision is due by next August. In the meantime, anti-testing activists say that the imminent EPA rule accepting human data would make such meetings commonplace.

A leaked draft of the regulations circulated in Washington last month. It would allow the EPA to consider human data from the industry provided the tests meet current ethical standards. Earlier studies would also be admissible as long as they met the ethical standards of their day.

“If the rule stays as this draft has proposed, the floodgates will open for human testing,” warns Erik Olson, a lawyer for the NRDC. ■

IN BRIEF

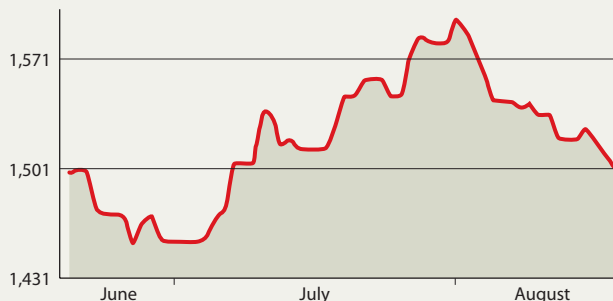
DESIGNER DODGE The US Food and Drug Administration last week approved a genetic test that identifies patients who are at increased risk of suffering side effects from a drug commonly used to treat colon cancer. The test, known as the Invader UGT1A1 Molecular Assay, is made by Third Wave Technologies in Madison, Wisconsin. Patients with some forms of the *UGT1A1* gene are not good at breaking down the colon-cancer drug irinotecan. In patients who test positive for one of these mutations, doctors can lower the drug dose to minimize the risk of any harmful reactions.

CLEAR VISION A biotechnology company whose fortunes rose and fell with its only product — a drug that stabilizes age-related vision loss — last week agreed to a takeover by OSI Pharmaceuticals of Melville, New York, for \$935 million. Eyetech Pharmaceuticals in New York City makes Macugen, a drug for people with age-related macular degeneration. The company's share price plunged this spring with word that Genentech's competing drug, Lucentis, which is still in clinical trials, not only stabilizes but also improves vision in people with one form of the condition.

OIL PRICES The largest state-owned Chinese oil company says that it has trumped an Indian competitor by striking a \$4.18-billion deal to buy PetroKazakhstan. This smallish oil company, based in Calgary, Canada, operates solely in oil-rich Kazakhstan, and its shares shot up 18%, to \$53.75, when China National Petroleum announced the purchase on 22 August. A rival joint venture in India — which includes ONGC Videsh, the international arm of India's state-controlled Oil & Natural Gas — has threatened to make a counteroffer.

MARKET WATCH

Nanotechnology stocks



Nanotechnology stocks moved steadily forwards earlier this summer as investors became a little less risk averse — only to slide back a bit in August.

The Lux Research index, which tracks companies with nanotech interests, has been following market trends for technology stocks over the past two months, says Peter Hebert, chairman of the New York-based consultancy that keeps the index.

Investors have leaned towards stocks in various high-tech sub-sectors since the spring, says Hebert, and nanotech companies have benefited from this trend. But, in line with other technology sectors, prices fell back last month.

Best-performing companies in the period, Hebert reports, included Veeco Instruments in Woodbury, New York, which supplies equipment such as atomic force microscopes. Its stock rose from \$16 to more than \$21 during July, on strong sales and earnings figures.

Another strong performer was

Westaim of Calgary, Canada, whose stock has risen by 40% since June on reports that it may spin off subsidiary Nucryst Pharmaceuticals, which has forged a successful business making nanocrystalline wound dressings.

It was a less sunny summer for Immunicon, a Pennsylvania specialist in cancer diagnostics using small magnetic particles; its stock fell 40% over uncertainty about its sales prospects.

And index participants whose fortunes are closely tied to the semiconductor industry — such as Oregon-based FEI Company, which sells equipment such as electron microscopes to chip-makers — were hurt by a weak capital-investment outlook in that sector.

The Lux index includes an array of specialized companies selling nanotech equipment or products, as well as some larger ones for whom the technology is important. ■

www.luxresearchinc.com

SOURCE: LUX RESEARCH

Time to use neuroscience findings in teacher training

SIR — Your Editorial “Bringing neuroscience to the classroom” (*Nature* **435**, 1138; 2005) and News Feature “Big plans for little brains” (*Nature* **435**, 1156–1158; 2005), on the emerging connections between education and cognitive neuroscience, are both hopeful and critical at once. There is currently a strong emphasis on the need for research findings to be both immediately available, and directly applicable to the classroom. This inadvertently sets high expectations which, if not met, could lead to the quick erosion of this developing field.

I believe it is time to think about the interactions between education and cognitive neuroscience in broader terms. In addition to conducting research projects such as those described in your Editorial and News Feature, it is now essential to begin integrating the teaching of scientific evidence from cognitive neuroscience into teacher-training and further-education programmes. This would facilitate the creation of a ‘researcher-practitioner’ model in the field of education. The US National Science Foundation centre at Boston University and the LearnLab project, as described in your News Feature, have made a start in this direction, but more of these types of projects are needed to create the basis for fruitful exchanges and collaborations between cognitive neuroscientists and educational practitioners.

The ‘science of learning’ needs to strengthen its conceptual backbone and put forth a broad set of aims, while permitting development to occur. If this is done, rejection on the basis of ‘lack of direct application’ can be avoided and rich outcomes anticipated.

Daniel Ansari

Department of Education,
Dartmouth College, Hanover,
New Hampshire 003755, USA

Embryo’s moral status is unaffected by alteration

SIR — Both proponents and opponents of embryonic stem-cell research should object to William Hurlbut’s proposal for nuclear transfer embryos to be genetically engineered to block their capacity for development into human babies (“Altered embryos offered as solution to stem-cell rift” *Nature* **436**, 309; 2005).

In describing such material as ‘embryo-like entities’, Hurlbut misses the point that that is what nuclear transfer embryos already are. Indeed, calling them ‘embryos’ seems somewhat tenuous, considering that they are

not the products of a sexual process; nor are they clones. But, whatever they are called, it is inescapable that any potential for development to babies can only be realized by implantation into the wall of a uterus. Engineering then seems pointless: in order to block the capacity to develop into a baby, simply don’t implant.

Deliberately altering genotype could have important consequences. Disablement of multitasking growth factors, for example, may interfere with cell-signalling mechanisms, thus affecting lab protocols for directed differentiation towards specialized cells in sufficient numbers for therapy.

From an ethical standpoint, intentionally downgrading the moral status of human embryos, in order to render them suitable for research that was otherwise deemed immoral, would be dissimulation.

We should be aiming to make this science more understandable and accessible to allow for proper informed debate. Instead, Hurlbut’s complicating proposal is self-contradictory and detrimental to the progression of important human embryonic stem-cell research.

Lee Turnpenny

Human Genetics Division,
University of Southampton,
Southampton SO16 6YD, UK

Industrial practices set no standard for science

Ian Taylor, in Correspondence (*Nature* **436**, 626; 2005), argues that a number of practices considered scientific misconduct in academia are acceptable to industry. In so doing, he may highlight the existing gulf between academia and industry, but in no way does he provide any support for the erosion of scientific principles, which should be defended to the utmost.

Applying the ideas of others may be commercially prudent, but it is not ‘science’. Building on current thought to produce new knowledge is science; simply applying it in a commercial setting is business. Similarly, it may be proper practice to withhold methodologies in an effort to protect proprietary information, but doing so reduces publications to the level of non-scientific anecdotal reports. If an experiment is published without full details of the methodologies, then not only is it impossible to reproduce the experiment, but the results, and indeed the methods, cannot be built upon by the wider scientific community.

Commercial and scientific interests may contribute one to the other, but their ends are different. I see no reason for academic research to adopt more commercial practices. Indeed, Taylor’s examples lend support to the opposite view — as well as to a proper

scepticism towards science conducted by commercial enterprises.

Steven Tait

Edinburgh Research and Innovation,
University of Edinburgh,
1-7 Roxburgh Street,
Edinburgh EH8 9TA, Scotland, UK

Farming need not replace fishing if stocks are rebuilt

SIR — In his Commentary article, “When will we tame the oceans?” (*Nature* **436**, 175–176; 2005), John Marra foresees mariculture as an important contributor to global food production and as a solution to overfishing. He argues that the world’s fisheries should be replaced by large-scale domestication of the oceans.

We agree that aquaculture is playing an increasingly important role in world fish supply. But a careful distinction must be made between the aquaculture of freshwater fishes, molluscs and plants — which is primarily low-tech and low-impact, and helps feed people in developing countries — and the high-tech mariculture of carnivorous finfish advocated by Marra, which serves luxury food markets. It is unlikely that low-income families will ever taste maricultured tuna, salmon or cod. Indeed, these people’s protein supply may diminish as the market for many small food fishes becomes cornered to provide mariculture fish feed.

Farming carnivores also results in a net loss of food because of inefficient energy conversion between trophic levels, as Marra acknowledges. Tuna farming, therefore, is not like herding cattle: it is the ecological equivalent of trapping wild shrews and foxes to feed caged wolves.

Fisheries and the wild populations that supply them should not be abandoned. Instead, serious effort should be focused on rebuilding depleted fish populations by creating large marine reserves and reducing total fishing capacity. Proper management of wild marine life could yield remarkable results, but—ironically—requires what Marra lists as a precondition for his vision of domesticated oceans: the political will to implement changes and create transnational agreements on shared ocean use. When we finally garner this political will, should we not use it to restore productive, biologically diverse ecosystems, rather than to risk further degrading the oceans?

Offshore mariculture is not “inevitable”. It is a course of action that can be chosen — or not.

**Julia K. Baum, Jana M. McPherson,
Ransom A. Myers**

Department of Biology,
Dalhousie University,
Halifax, Nova Scotia B3H 4J1, Canada

COMMENTARY

The ethics of research on great apes

In the wake of the chimpanzee genome publication, **Pascal Gagneux, James J. Moore and Ajit Varki** consider the ethical and scientific challenges for scientists who work on captive great apes.



Most captive great apes were born in captivity and returning them to the wild is not feasible.

J. McDONALD/CORBIS

Publication of the draft sequence of the chimpanzee genome is an exciting event; it opens the door to learning a great deal about our closest evolutionary cousins — and about ourselves in the process. But unlike the human genome project, the chimpanzee sequencing effort was not accompanied by studies addressing ethical, legal and social issues¹. Meanwhile, there is continuing debate over the future of captive ‘great apes’ (chimpanzees, bonobos, gorillas and orang-utans)*.

What does the publication of the chimpanzee genome mean for the thousands of great apes in captivity in the United States? Some fear the potential for increased invasive research on these individuals. Others are concerned that our limited knowledge of chimpanzee physiology and biology will constrain the usefulness of the chimpanzee sequence for understanding both humans and great apes.

For example, critical resources required for comparative genetic and biological studies, such as messenger RNA or complementary DNA libraries, are almost non-existent for great apes. Here, we advance a proposal that addresses these and related issues, to lead, we hope, to a mutually beneficial outcome for all, including the great apes (see Box for a summary of proposed goals and objectives). We emphasize that this article relates only to great apes, and not to other primates, nor other animals. Also, this piece is not about animal ‘rights’ but about ethical and scientific challenges specific to great apes in captivity.

Born in captivity

Opinions and attitudes regarding captive great apes span from the view that they are just expensive research animals to the idea that they should be accorded equal ‘rights’ with humans.

Such views are in the minority, but there is need for continued dialogue among the majority spanning the middle ground.

The current ethical status of the great apes also varies among nations. US research on great apes is regulated by local ‘animal-subjects’ committees. And although national guidelines for breeding and long-term care have been proposed^{2,3}, there is still much disagreement. Some believe that our close similarity to the great apes means that they should never be kept in captivity, but for the ones now living in US facilities, it is too late.

While great ape numbers in the wild have fallen to tens of thousands, captive populations have expanded, especially in the United States, where past government support for breeding programmes was aimed at producing subjects for research into the human immunodeficiency virus (HIV). Today the United States is

*Footnote: ‘Great apes’ is used here in its colloquial sense. In the commonly used classification, these species are grouped alongside humans in the family Hominidae, and humans belong to the tribe Hominini, along with chimpanzees and bonobos.

home to roughly 3,000 captive great apes (mostly west African chimpanzees) in research institutions, sanctuaries, zoos, private hands or the entertainment industry. Most of these individuals were born in captivity and never learned how to forage for survival or avoid predators. Thus, with few exceptions, attempts at returning captive great apes to the wild have proven extremely demanding — logistically and financially.

Regardless, we agree with those who say the biomedical-research community has special ethical responsibilities towards captive great apes. In our view, the great apes share traits — including, but not limited to, their genetic similarity to humans, the ability to use and modify tools and a sense of 'self' — that collectively justify this special status. (Individually, such traits are not unique to great apes; for example, bottle-nosed dolphins may also have a sense of self.)

Pause for thought

But there are other reasons to re-evaluate the situation for captive great apes. Their current medical care often assumes physiological and pathological identity with humans. But despite genetic and biological similarities, humans and apes differ markedly in their susceptibility to some major diseases, including AIDS (ref. 3). Working out the reasons for such biomedical differences will benefit all concerned, including the great apes, by allowing more species-appropriate medical care. Understanding how our genetic differences give rise to these and other biological differences has been a long-term interest of some researchers.

Sequencing of the chimpanzee genome is likely to motivate many further studies of ape biology and physiology. But how such research should proceed needs careful thought. Given the diversity of opinions (including among the three of us), it is impossible to define a single

clear-cut principle that can guide this discussion. We do suggest, however, that the study of great apes should follow ethical principles generally similar to those currently used in studies on human subjects who cannot give informed consent. Of course, many complex questions arise, such as who acts as the advocate for a great ape in agreeing on what are appropriate studies? And there are many grey areas. For example, is it acceptable to do reversible harm, such as causing a mild treatable infection (as is done with adult human volunteers), or to sedate a chimpanzee (as you would a child) to allow a therapeutic or research procedure?

Captive great apes have been subject to experimental procedures with the potential for irreversible damage or death, such as infections with human pathogens, vital-organ biopsies, multiple inoculations for vaccine testing, transfections for virus production and so on. Development of the widely used hepatitis B vaccine and understanding of the hepatitis C virus would not have been possible without the use of captive chimpanzees — and may still not be possible using other technologies. In retrospect, however, many of these expensive studies (for example, on HIV/AIDS, *Plasmodium falciparum* malaria and influenza A) turned out to have limited benefits for improving human health.

We suggest that alternatives to the use of whole chimpanzees be sought as soon as possible, and that substantial new funding be directed towards finding such alternatives. And, as with humans, we believe that the newly emerging genomic data should never be used to attempt germline genetic modifications in great apes (to produce 'transgenic' apes, as is routinely done with mice). Additionally, we recommend that any new

biomedical studies on great apes be carried out in a manner that supports further improvements to their care.

The time has come to establish broadly accepted guidelines for systematic, humane and ethical studies of captive great ape populations. These studies should be carried out at all levels, from genetics to biochemistry to physiology to behaviour and culture. A previous US National Research Council report²

addressed many issues regarding the care of captive chimpanzees, and a follow-up 2005 Federal Register Notice emphasized that they deserve the best and most humane care possible. For example, they should be maintained in groups that respect existing social bonds, with opportunities for physical, intellectual and social activities. Moreover, euthanasia is specifically excluded as a means of population control². Although opinions vary about the benefits of contact with human caretakers, there is generally wider agreement regarding human intervention for the control of escalating aggression within or between groups.

Precious resource

There is currently a moratorium on the breeding of chimpanzees at facilities funded by the National Institutes of Health (NIH). Although this may seem inhumane to some, it must be remembered that each birth in captivity can represent a 50-year or longer commitment on the part of human society. Facilities that do allow great apes to breed should avoid large numbers of births, as well as inbreeding and the mixing of subspecies.

As long as great ape facilities provide a safe, healthy and humane environment, it seems reasonable that captive great apes should remain a source of basic knowledge — which, in turn, may benefit both them and us. Understanding the normal biology, physiology and behaviour of the great apes provides a unique approach to understanding ourselves, even if we do not suffer from all the same diseases. Much of this can be accomplished through simple observational studies and by giving high-quality medical care to diseased individuals, as occurs routinely in human medicine. Experiments involving physical intervention with no long-term consequences could also be considered, provided that there is due consideration to the individual personalities of each ape, and that comparisons to normal humans are made wherever possible.

When a captive ape dies of natural causes (or is humanely killed to end incurable suffering), a thorough autopsy and rapid collection of organ samples for genomic, transcriptomic (gene expression), proteomic, biochemical and histological studies should be done, to generate an extremely valuable and sorely needed resource. There is also much to learn by careful preservation and analysis of the

A summary of proposed goals and objectives

Community issues

- Promote funding for an ELSI (ethical, legal and social issues) component of the chimpanzee genome project, as was done with the human genome project.
- Encourage dialogue on ethical standards and guidelines for research on great apes, following principles generally similar to those used in research on humans.
- Promote institutional and individual recognition of, and support for, the connection between the care and use of captive apes and their conservation in the wild.

Research issues

- Encourage exploration of genetic, biological and medical similarities and differences between great apes and humans, especially in the context of providing medical care.
- Promote development of standardized databases of individual genotypic and phenotypic information about all captive great apes.

- Encourage funding for standardized collection and banking of tissues, fluids, imaging and biometric data obtained during medical care and autopsies. And make such data available to the scientific community for genetic, biochemical, histological and morphological studies.
- Encourage funding for the production of high quality cDNA libraries.
- Encourage funding for expanded programmes focused on understanding cognitive functions in great apes.
- Encourage development of mechanisms for sharing data, while respecting individual and institutional privacy concerns.

Care issues

- Encourage greater fiscal support to ensure optimal living conditions for captive great apes.
- Suggest mechanisms to ensure and support the best possible medical care for captive great apes.

remaining musculoskeletal system. Partly due to inadequate funding, personnel, and facilities, many great ape deaths now occur without such analysis, translating into numerous wasted opportunities to learn more about their biology. Being responsible for great ape captivity, we must maximize the information from them, rather than treating them as single-use, disposable tools. Likewise, body fluid and tissue samples that are collected during routine medical care are often discarded or inadequately archived. Such detailed studies of living and deceased humans have long benefited our species by providing valuable medical and scientific knowledge. Indeed, some humans approve postmortem donation of their entire bodies to science.

Mutual gains

In 2000, the US Congress passed a Chimpanzee Health Improvement, Maintenance, and Protection Act mandating the establishment of the NIH Chimpanzee Management Program⁴ (ChiMP) and federal funding of sanctuaries for chimpanzees from research institutions, such as Chimp Haven (www.chimphaven.org). We suggest that these and many other ongoing efforts be bolstered by a federally and philanthropically supported collaborative network in which facilities housing captive great apes could choose to participate.

This would generate interactions among interested scientists from fields such as comparative biomedicine, psychology or biological anthropology. Already, leaders from US institutions holding most chimpanzees have come together to establish a National Chimpanzee Resource Committee, which meets regularly to discuss issues of mutual interest. The increased cost of supporting all such facilities will be more than justified by the knowledge gleaned from the study of healthy, socially integrated great apes — information that could potentially contribute to the ultimate survival of some of these species in their natural habitat.

Such a national network could also help train and support scientists interested in the standardized accumulation of all relevant biological information on healthy captive great apes. Each great ape should continue to be accounted for, by a name and unique identifier. Complete medical records should be collected in a standardized fashion into electronically searchable databases, in a way that maintains the privacy of researchers and institutions. Samples, such as body fluids, taken from live apes during routine physical examinations should also be collected and archived. In this way, we can create a great ape tissue bank of flash-frozen and archived samples for use by the scientific community — which could eventually result in (among other payoffs) the production of high quality cDNA libraries.

In some cases, therapeutic medical care could be extended to include data collection



J. BALOG/GETTY

Fast learner: tool use is one of the traits that sets the great apes apart from most other research animals.

for research purposes (for example, standardized brain magnetic resonance imaging protocols appended to diagnostic imaging procedures). Increased funding will be needed to enhance existing medical facilities and expertise, and the ability to perform complete autopsies with tissue collection.

As for newly proposed research studies on live great apes, we suggest that these be reviewed and approved by specialized ethical oversight groups that incorporate appropriate aspects of the separate human-subject and animal-subject committees found at most institutions. Cooperation by the great ape research subjects will be critical for many studies, and will only be possible if there is also adequate funding for behavioural training of the animals.

We fully recognize that our proposal is unlikely to please everyone interested in great apes, and that this is only an initial contribution to a much-needed dialogue among all interested parties. Many changes and adjustments will be required to develop a mutually acceptable solution for all concerned, including the great apes.

Meanwhile, there is a deep irony in the fact that the sequencing of the chimpanzee genome coincides with the potential demise of great apes in the wild. We urge all scientists studying great apes, or tissues and samples derived from them, to contribute not only to the care of captive apes, but also to develop mechanisms by which studies of captive great

apes would help generate a revenue stream to support the conservation of populations in the wild. While recommending improved care of captive great apes, we recognize that the remaining wild great apes may end up living in strictly managed reserves, depending on increased human intervention for their survival. In the long run, even our ability to care for wild populations could benefit from an increased understanding of great ape cognition, behaviour, physiology, biology, pathology and medicine.

Pascal Gagneux is at Conservation and Research for Endangered Species, Zoological Society of San Diego, San Diego, California, USA; James J. Moore is in the Department of Anthropology, and Ajit Varki is in the Department of Medicine and Cellular & Molecular Medicine, University of California, San Diego, La Jolla, California, USA.

1. Human Genome Project Information www.ornl.gov/sci/techresources/Human_Genome/elsi/elsi.shtml (2004).
2. The National Academies Press <http://books.nap.edu/catalog/5843.html> (1997).
3. Olson, M. & Varki A. *Nature Rev. Genet.* **4**, 20–28 (2003).
4. National Center for Research Resources www.ncrr.nih.gov/compmed/cm_chimp.asp

Acknowledgements: We are grateful to the following readers for very helpful comments and suggestions: A. Zihlman, C. Tutin, D. Povinelli, D. Rumbaugh, F. B. M. de Waal, J. Goodall, J. Allman, K. Semendeferi, K. Benirschke, M. Goodman, O. Ryder, R. Wrangham, S. Boysen, S. Blaffer Hrdy, S. Savage-Rumbaugh, T. Matsuzawa, T. Murray, and W. McGrew. We also gratefully acknowledge the support of the G. Harold and Leila Y. Mathers Charitable Foundation.

COMMENTARY

A unique biomedical resource at risk

Research using chimpanzees has been crucial in the fight against human diseases such as hepatitis.

John L. VandeBerg, Stuart M. Zola and colleagues urge that this now dwindling resource be sustained.

As the closest living relative of humans, the chimpanzee holds a unique place in biomedical research. Several major medical advances have been possible only through research with chimpanzees. The completion of the draft sequence of the chimpanzee genome, the first non-human primate genome to be sequenced, places the chimpanzee in a position to play an increasingly critical and invaluable role in future biomedical advances. But with the decline in the US population of captive chimpanzees available for research, this national resource is at risk. Lifting the current moratorium on breeding chimpanzees in US research facilities is essential to preserve this unique resource — which we believe is needed to solve some of the most important global health problems of today and tomorrow.

Top model

Research with chimpanzees is essential for reducing risks to human research subjects, and ultimately to human patients. For example, the need for chimpanzees to test monoclonal-antibody therapies is rapidly increasing as more antibodies are designed. Monoclonal antibodies, genetically engineered to be almost identical to human antibodies, hold great promise for treating cancer; autoimmune diseases, such as rheumatoid arthritis, lupus erythematosus, multiple sclerosis, psoriasis, and Crohn's disease; asthma, poisoning, and septicemia; and virtually any disease caused by viral infection^{1,2}.

Therapeutic antibodies are produced from a single (mono) hybridoma cell line derived from one antibody-secreting spleen cell of a mouse and one myeloma cell. Many of the biological receptors targeted by these antibodies are present only on chimpanzee and human cells. Moreover, in monkeys and other mammals, human-like monoclonal antibodies are rapidly cleared from the bloodstream, so that the safety and efficacy of repeated doses cannot be tested. Because the antibodies do not elicit immune responses in chimpanzees, they persist in the blood as they do in humans, and their effects can be evaluated.

At least 11 different monoclonal-antibody therapies have already achieved Food and



Star player: Clint, who died in January, provided the DNA used to sequence the chimp genome.

Drug Administration (FDA) approval, and more than 400 others are in clinical trials, including 70 that have proceeded to phase-II testing or beyond^{3,4}. Some of these antibodies were tested in chimpanzees before they entered clinical trials (proprietary data, unpublished). Data for antibodies proposed for clinical trials (but not data for antibodies that produced side-effects or were ineffective in chimpanzees), are supplied to the FDA. However, these data are not published.

It has also recently been demonstrated that chimpanzees predict human pharmacokinetics — the time course of absorption, distribution, metabolism and excretion of drugs in the body — more accurately than other animal models, including rats, dogs and other non-human primates⁵. Before it enters clinical trials with human subjects, every drug must undergo extensive pharmacokinetic characterization to determine the appropriate dosage for the drug to be effective, but not toxic.

All drugs are initially tested in non-primate animal models for toxicity and pharmacokinetic characteristics. However, sometimes the results from several species are inconsistent; in such cases, chimpanzees are required for the final pre-clinical testing. As the number of new drugs being developed increases, the need to use chimpanzees in pharmacokinetic evaluations will continue to rise. Most pharmacokinetic evaluations require only a few days, so a single chimpanzee can be used for testing many drugs each year. Therefore, a limited population of chimpanzees will suffice for testing a large number of drugs.

Next wave

Armed with the chimpanzee genome sequence, in conjunction with the sequence of the human genome, research on chimpanzees could also help us to identify genetic and phenotypic specializations that are relevant to human disease, and ultimately new therapies.

*Footnote: The authors are members of the US National Chimpanzee Resource Committee (NCRC) and include representatives of each of the major centres that maintain the national chimpanzee research resource.

For example, certain cancers seem to occur at much higher rates in humans than in other primates and humans are uniquely vulnerable to some of the pathological correlates of neurodegenerative diseases, for example Alzheimer's disease⁶. Because human specializations are, by definition, features of humans that evolved after the separation of the human and chimpanzee lineages, the identification of such specializations is possible only by comparing humans with chimpanzees.

Chimpanzees continue to play a vital role in our efforts to discover and understand the mechanisms that underlie a wide range of infectious diseases, including hepatitis — a constellation of diseases that cause liver inflammation. The chimpanzee is the only animal model that can be successfully infected with the hepatitis B and C viruses, and is therefore critical to research aimed at developing vaccines, and drugs to prevent and treat hepatitis B and C. Indeed, a safe and effective vaccine against hepatitis B infection, which has been available since 1982, was developed from research with chimpanzees.

Deadly threats

Although this vaccine is quite effective, better drugs for treating chronic hepatitis B infections are badly needed. Moreover, about 170 million people, including 2.7 million in the United States, are chronically infected with hepatitis C virus (HCV); 70% of them will develop chronic liver disease (see www.who.int/mediacentre/factsheets/fs164/en and www.cdc.gov/ncidod/diseases/hepatitis/c/fact.htm). Liver failure caused by HCV infection is the leading reason for liver transplants. No effective vaccine has yet been developed for hepatitis C, and existing therapeutic drugs have limited efficacy in preventing progression of liver disease.

The discovery that malaria is far more prevalent than anyone had realized has also generated major concern in public-health centres around the world⁷. In 2002, more than 500 million people — nearly double the estimate in 1995 — were infected by the deadliest form of malaria. Of the four malaria parasites, *Plasmodium falciparum* is by far the most dangerous to humans, especially to the undernourished, weak or very young. It is prevalent throughout the tropics and has developed resistance to several drugs.

Chimpanzees have their own set of malaria parasites, including *P. reichenowi* — *P. falciparum*'s direct counterpart and closest relative. Differences between the human and chimpanzee *Plasmodium* species provide an opportunity to better understand the biological characteristics of the human parasite and, perhaps, to develop interventions against it. Because *P. reichenowi* will not infect humans or any other non-human primate species and

"The scientific value of chimpanzees is certain to increase as a consequence of the completion of the genome sequence."

Table 1 | Five-year projection of the number of chimpanzees available for research in the absence of breeding.

Biomedical research centre		Number of chimpanzees available for research as of 1 April 2005 (total = 1,171)	
New Iberia Research Center		373	
Alamogordo Primate Facility		247	
Southwest National Primate Research Center		236	
M.D. Anderson Cancer Center		133	
Yerkes National Primate Research Center		109	
Primate Foundation of Arizona		73	
Date	Projected reductions		Number of available chimpanzees
	Reason	Number	
1 April 2005			1,171
	Mortality ($1,171 \times 0.03 \times 0.75$ yr)	-26	
1 January 2006			1,145
	Removal to Chimp Haven	-174 ^a	
	Mortality ($(1,145 - 174) \times 0.03$)	-29	
1 January 2007			942
	Removal to Chimp Haven	-20	
	Mortality (922×0.03)	-28	
1 January 2008			894
	Removal to Chimp Haven	-20	
	Mortality (874×0.03)	-26	
1 January 2009			848
	Removal to Chimp Haven	-20	
	Mortality (828×0.03)	-25	
1 January 2010			803
	Removal to Chimp Haven	-5	
	Mortality ($798 \times 0.03 \times 0.25$ yr)	-6	
1 April 2010			792

These projected reductions assume that: no chimpanzees are produced; the annual mortality rate of the population at the research centres is 3%; and that Chimp Haven, the national sanctuary, will be filled and capable of housing 200 chimpanzees by the end of 2006, including the 31 currently housed there. ^aThis figure includes 169 to maximize the population at Chimp Haven, plus five to replace projected deaths among the current 31 animals during 1.75 years (from 1 April 2005 to 31 December 2006). The annual mortality rate at Chimp Haven is 10% because most of the animals arriving at the facility are old.

has not been successfully cultured *in vitro*, the only means we currently have for propagating the parasite is to use chimpanzee hosts to develop large populations of parasites, which are harvested by periodic blood collections.

In 1997, the National Research Council Committee on Long-term Care of Chimpanzees, funded by the National Institutes of Health (NIH), published a landmark report⁸. Two of the committee's recommendations have had a profound impact on the national chimpanzee resource. These were that "a breeding moratorium should be imposed for at least five years (1997–2001)" and that "sanctuaries capable of providing for the long-term care and well-being of chimpanzees that are no longer needed for research and breeding should be established."

It is important to note that these recommendations were made following what seemed to have been an overproduction of chimpanzees in response to the discovery that chimpanzees were the only animal that could be infected with HIV-1; it was believed that large numbers of chimpanzees would be

required for fundamental research on AIDS and for testing potential AIDS vaccines.

Fast decline

Indeed, during the 1980s and 1990s, chimpanzees played a critical role in clarifying our basic understanding of HIV-1, and in the testing of potential vaccines⁹. However, few HIV-1-infected chimpanzees progressed to a state of immunodeficiency. And the usefulness of chimpanzees in studying the progression of AIDS was eventually displaced by work with monkeys, in which HIV-1-like viruses produce infections and clinical signs resembling those in humans suffering from AIDS. Nevertheless, chimpanzees are the only natural animal model that can be infected with HIV-1. Thus, they are still important for testing vaccines aimed at preventing HIV-1 infection or reducing the virus load in infected individuals.

The breeding moratorium on NIH-owned or NIH-supported chimpanzees continues today. In general, institutions without NIH funding have also chosen to discontinue breeding because of concerns that federal and commercial support for research with chimpanzees may not cover the lifetime expense of maintaining the animals, which can live beyond 40 years of

age. In addition, several hundred chimpanzees that are no longer useful for biomedical research have been or soon will be transferred to sanctuaries, where they will not be available for biomedical research or breeding¹⁰. These two actions, together with natural attrition by death in the ageing chimpanzee population, have led to a dramatic reduction in the number of chimpanzees that are available for research and breeding purposes (see Table 1).

In 1997, the number of chimpanzees estimated to be available for breeding or for biomedical research was 1,494 (ref. 8). On 1 April 2005, the census of such chimpanzees was 1,171 — a 22% decline in the overall population (see Table 1 for numbers of chimpanzees available at biomedical research centres as of this date). Moreover, the number of available chimpanzees is expected to continue to decline rapidly. Only 792 chimpanzees are predicted to be available in five years, a 47% decline since 1997 (see Table 1).

A global problem

The 1997 National Research Council⁸ report also recommended that “a core population of approximately 1,000 chimpanzees be assured lifetime support by the federal government.” Already, the number of chimpanzees supported by the federal government is well below 1,000 (R. O'Neill, personal communication), and a significant proportion of the projected 792 existing chimpanzees expected to be available in five years will be privately owned and supported. So although the federal government implemented recommendations for a breeding moratorium (and extended it beyond the initial five-year time-frame), and for retiring research chimpanzees that are no longer useful to sanctuaries, it has not implemented the recommendation to ensure a federally supported core population of 1,000 chimpanzees.

Because biomedical research with chimpanzees has been terminated in most other countries (see ref. 10), the animals in our national facilities are a global resource. We believe that the decisions to terminate the experimental use of chimpanzees in other countries were not based on scientific grounds, but mostly on financial considerations or on philosophical grounds. Since those decisions have been made, there has been an increasing frequency of enquiries from scientists and commercial companies in Europe for access to US chimpanzees. We believe that the 6.3 billion people in the world today, as well as those yet to be born, will depend on the small number of chimpanzees available for research for some of the most dramatic medical advances of the future.

Many advances from biomedical research with chimpanzees have been published in the past one to two years, demonstrating that



With no breeding, the ageing population of chimps in US research facilities will eventually disappear.

rapid medical progress pertinent to a wide range of human diseases is being made through the use of chimpanzees. Indeed, both the scientific value of chimpanzees and their use in biomedical research are certain to increase as a consequence of the completion of the chimpanzee genome sequence, the development of powerful microarray and proteomic technologies, and the establishment of new capabilities in immunology enabled by monoclonal-antibody technologies.

Preparing for the future

Although we have suggested some potential future uses for chimpanzees in biomedical research, history implies that we will almost certainly not have predicted some of the most important ones. Had we written this manuscript 35 years ago, we could not have predicted the important role chimpanzees would play in understanding the basis of AIDS and the difficulties in producing vaccines for it; AIDS was not known. We could not have predicted that chimpanzees would be the key to developing antiviral drugs and vaccines for hepatitis A, B and C, because the viruses that

“The window of time in which to establish a renewable resource of chimpanzees has become dangerously short.”

cause these three forms of hepatitis had not yet been identified. We could not have predicted the critical role that chimpanzees would play in the development of antibody technologies to treat cancer, autoimmune diseases, and other human conditions, as monoclonal antibodies had not even been conceived. And RNA-interference technology, which will require chimpanzees for its development as a therapy for hepatitis and many other diseases, was discov-

ered only during the past decade.

We know that the chimpanzee genome sequence will enormously increase the scientific value of chimpanzees in the diverse areas of biomedical research for which they are currently used. However, we cannot predict which of the human epidemics that will emerge during future decades will require chimpanzees for the development of preventions and treatments, and we cannot predict what powerful new technologies will emerge that can only be fully exploited to prevent and treat human diseases if a robust national chimpanzee resource is available.

What we can predict with confidence is that new epidemics will arise, as they have done throughout history, and that technological developments that we cannot imagine today will accelerate even more rapidly than they have done in the immediate past. The chimpanzee is certain to remain a unique animal model for research on the prevention and treatment of a multitude of human diseases, and we predict that it will become even more valuable in the future than it has been until now.

To meet these global challenges, it will be necessary to establish a renewable and robust national resource of chimpanzees in sufficient numbers to meet the needs of biomedical research. The window of time during which this objective can be accomplished has become dangerously short as a consequence of the breeding moratorium that was imposed eight years ago. We call upon the leaders of the NIH to fulfill this national, and global, need by assuring the availability of chimpanzees for biomedical research. ■

John L. VandeBerg is at the Southwest National Primate Research Center, San Antonio, Texas 78245, USA, and Stuart M. Zola is at the Yerkes National Primate Research Center, Atlanta, Georgia 30329, USA.

Co-authors are Jo Fritz who is at the Primate Foundation of Arizona, Mesa; D. Rick Lee who is at the Alamogordo Primate Facility, New Mexico; Thomas J. Rowell of the University of Louisiana at Lafayette, Louisiana; William C. Satterfield of the M. D. Anderson Cancer Center, Bastrop, Texas.

1. Stern M. & Herrmann, R. *Crit. Rev. Oncol. Hematol.* **54**, 11–29 (2005).
2. Berger, M., Shankar, V. & Vafai, A. *Am. J. Med. Sci.* **324**, 14–30 (2002).
3. Desgranges, C. *Pathol. Biol. (Paris)* **52**, 351–364 (2004).
4. Smith, K. A. *et al. J. Clin. Pathol.* **57**, 912–917 (2004).
5. Wong, H. *et al. Drug Metab. Dispos.* **32**, 1359–1369 (2004).
6. Gearing, M. *et al. Proc. Natl Acad. Sci. USA* **91**, 9382–9386 (1994).
7. Snow, R. W., Guerra, C. A., Noor, A. M., Myint, H. Y. & Hay, S. I. *Nature* **434**, 214–217 (2005).
8. National Research Council *Chimpanzees in Research: Strategies for their Ethical Care, Management and Use* (Natl Acad., Washington DC, 1997).
9. Murthy, K. K. *et al. AIDS Res. Hum. Retroviruses* **14** (suppl. 3), S271–S276 (1998).
10. Brent, L. *Lab. Anim.* **33**, 37–43 (2004).

Acknowledgements: We thank J. Barnwell, M. Galinski, R. E. Lanford, T. M. Preuss, R. H. Purcell, C. Walker, and S. Williams-Blangero for their suggestions and contributions to this manuscript. We also acknowledge support of our chimpanzee resources by NIH resource and research grants.

BOOKS & ARTS

A family resemblance

We are closely related to other apes, but how similar are we really?

Our Inner Ape

by Frans de Waal

Riverhead: 2005. 288 pp. \$25.95

To be published in the UK in November
by Granta.

Robert Sapolsky

Scientists often become agitated when confronted with someone who does not believe in evolution — especially when, as in the United States, such people try to dictate what facts our schoolchildren are taught. Only slightly less distressing is meeting someone who enthusiastically embraces evolution but with enough misconceptions to make Darwin turn in his grave. In this excellent book for the public, Frans de Waal tackles some exasperating misconceptions about the evolution of the social behaviour of apes, particularly humans.

The misconceptions rest on two things: Jane Goodall and the number 98. Goodall has long astonished the world with her discoveries about the behaviour of wild chimpanzees. And it was she who reported the profoundly disturbing fact that chimpanzee social life can include murder, cannibalism and organized inter-group violence, findings widely disseminated by the *National Geographic*. And 98 is approximately the unnervingly high percentage of DNA that humans share with chimps. Combine those two and our fate appears sealed: our closest relative, with whom we are nearly genetically identical, is a murderous thug. There you have it, our human-as-killer-ape destiny.

The antidote to all this is the bonobo. Once known as the pygmy chimp, the bonobo is now recognized to be a separate species, and from taxonomic and genetic standpoints, we are as closely related to it as to the chimpanzee. And the bonobo is very different. Males are not particularly aggressive, and lack the massive musculature typical of species (such as chimps) in which a male's ability to pass on copies of his genes depends heavily on his ability to pummel other males. Moreover, the bonobo social system is female dominated, food is often shared, and there are well developed means for reconciling social tensions.

And then there is sex. Bonobo sex is the prurient highlight of primatology conferences and it makes parents shield children's eyes when watching nature films. Bonobos have sex in every conceivable (and inconceivable)

Despite appearances, humans have come a long way since we branched off from the other apes.

position, in pairs or otherwise, within gender, between gender, to greet someone, to solve social conflicts, to work off steam after being scared by a predator, to celebrate finding food, to cajole the sharing of it... or just because. As has become the sound bite to contrast the two species, chimps are from Mars and bonobos are from Venus.

De Waal is uniquely qualified to be our guide to the different social worlds of these two relatives, as an immensely accomplished primatologist and an expert on both species. For example, his book *Chimpanzee Politics* (Jonathan Cape, 1982) is a classic analysis of chimp machinations over power, and *Bonobo* (University of California Press, 1997) is one of the few monographs on the species.

Our Inner Ape is organized into chapters analysing the two species' use of power, their sexual and aggressive behaviours, and their capacity for kindness. What's more, the writing is clear and witty.

So, one asks breathlessly, is de Waal's line that we humans are like chimps or bonobos? A lesser scientist with one ideological bent would say this isn't a perfect world and surrender to our unfortunate chimpiness. And an ideologue

of a different stripe would debunk the idea of our chimpish nature and search for our inner bonobo. But de Waal raises deeper points.

First, neither chimps nor bonobos are what humans were like in our ancestral past. While we've been busy evolving in the 5 million or so years since the last ancestor we shared with these species, chimps and bonobos have not been frozen museum displays: they are our contemporaneous cousins, not our ancestors.

Second, in every realm of our behaviour and biology, we humans bear some striking similarities, and differences, to each of the species. Studying another species in order to gain an understanding of our own is as much about the differences as the similarities, observing the unique solutions that each species has come up with to solve its own evolutionary, ecological and social challenges.

Finally, the dramatic dichotomy between those fun 'party animal' bonobos and those mean old chimps is somewhat exaggerated. Bonobos still have hierarchies and conflict (why else have reconciliation?); their peaceful kingdom is not built on inherent egalitarianism, but on necessary tolerance, and bonobos have most plausibly developed that frothy

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

sexuality as a means of avoiding infanticide. Meanwhile, chimps too are capable of affiliation and altruism; their violence and competition are embedded in an array of social checks and balances, and they even display something resembling empathy.

De Waal covers this with great wisdom and subtlety. I have problems with his views in only one domain. In considering that even bonobos are hierarchical, de Waal suggests that hierarchies are inevitable for social species — “we simply could not live without them” — for a number of reasons. First, if we lacked easy ways to discern rank, “no one would be able to tell who is who”. A response to this is that when rank and merit dissociate, a lack of easy sym-

bols for the former makes it easier to recognize and value the latter. Second, de Waal sees rank as allowing for grander human accomplishments. “This is why the most cooperative human enterprises, such as large corporations and the military, have the best-defined hierarchies.” But this assumes that corporations or armies are truly cooperative enterprises, as opposed to subordinating systems that benefit profiteers (such as shareholders, or leaders who control armies). Finally, de Waal sees an ironic and paradoxical benefit to hierarchy, as it can take an abusive hierarchy to provide the impetus for cooperation from below to overthrow it and establish something fairer. Naturally, one hopes instead for hunter-gatherers

writ large, where justice emerges naturally from bottom-up local interactions, rather than “democracy as born from violence”, to use de Waal’s phrase. But, as John Lennon put it, you may say I’m a dreamer, which is quite possible.

That difference of opinion aside, this is a rarity, a superb scientist producing an excellent book for non-specialists. This should be required reading for the opinionated cousins (or better yet, world leaders) whose ancient encounters with Robert Ardrey or Konrad Lorenz have led them to believe that they understand what kind of ape we are. ■

Robert Sapolsky is in the Department of Biological Sciences, Stanford University, Stanford, California 94035, USA.

Symmetry by numbers

The Equation that Couldn't be Solved: How Mathematical Genius Discovered the Language of Symmetry

by Mario Livio

Simon & Schuster: 2005. 368 pp. \$26.95

István Hargittai

The equation in the title is the quintic equation, the mathematical genius is Évariste Galois (1811–32), and the language of symmetry he discovered is group theory. Symmetry combines both beauty and science, and can easily be seen in the world around us. But before he could use it in science, Galois had to create the necessary mathematical tools. The world was slow to listen, and it took almost a hundred years for the practical value of group theory to be truly appreciated. Galois, meanwhile, was killed in a duel at an early age. In *The Equation that Couldn't be Solved*, Mario Livio follows his brief existence like a sleuth.

Born into a scholarly family in a Paris suburb in Napoleonic France, Galois was educated at home before being sent to a boarding school in Paris that rivalled the English schools of the time for austerity and rigid discipline. He was not a great success at school, but soon found satisfaction in mathematics, which became his sole occupation by the time he was 16. Having failed to gain entrance to a more prestigious college, he continued his studies in a high school.

Galois was still only 17 when he continued work started by the Norwegian mathematician Niels Henrik Abel, showing in general terms whether an equation is solvable by a formula or not. For this he introduced the seminal concept of a group, and created a new branch of algebra now known as Galois theory. His first publication appeared in 1829, but a combination of neglect and egotism prevented senior mathematicians of the day from giving him the exposure he deserved. When the work was finally introduced to the French Academy of Sciences, it was hardly appreciated.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Duel publication? Évariste Galois scribbled down his notes on group theory on the eve of his death.

Nonetheless, Galois continued his creative work, against all the odds. He failed another entrance examination as people greatly inferior to him could not appreciate his work, and lost his adored father, a Republican who was driven to suicide by his royalist political opponents. Young Galois also had a passion for Republican revolution and served a prison term for his political activities. He fell in love with an undeserving girl and was killed in a duel that was related to this unfortunate entanglement. During the night before the tragedy, Galois hurriedly wrote a profound description of his group theory, remarking in the margin: “I have no time.”

The Equation that Couldn't be Solved covers a remarkable number of different topics, including biographies of scientists and mathematicians. It also covers the Rubik cube and other puzzles; string theory; supersymmetry; the origin of creativity; the relationship between the external symmetry of the human face

and body, and mate selection and sex life; and much more. Livio examines the contributions of others that led up to Galois’ discovery, and gives a panoramic view of the direct, as well as quite remote, applications of group theory.

Very little escapes Livio’s attention, especially in twentieth-century physics. But one omission is the contribution to the story of Eugene Wigner. He applied group theory to quantum mechanics in the 1920s, when most of his contemporaries were yet to value it: Wolfgang Pauli called it “die Gruppenpest” — roughly translated as “that pesky group business”. Wigner was awarded a Nobel Prize in 1963 for this work.

Another omission from the book is that, in discussing crystallography, Livio stops at the classical notions of symmetry and defines crystallography as “the science studying the structures and properties of assemblies made of very large numbers of identical units”. This idea supposes regularity and periodicity, and was largely a result of the tremendous success of X-ray diffraction in the twentieth century. Recently, however, the field has embraced other structures, such as the newly discovered quasicrystals of regular but non-periodic patterns. It was an early suggestion by British crystallographer Alan Mackay that the rules describing ‘crystal’ structures be relaxed — and they have been. They now include structures that fall beyond the 230 space groups, and the new rules do not necessarily form groups.

Overextending the inferences from symmetry can be restrictive. As the historian of mathematics E. T. Bell said: “The cowboys have a way of trussing up a steer or a pugnacious bronco, which fixes the brute so that it can neither move nor think. This is the hog-tie and it is what Euclid did to geometry.”

The book seems a little biased in places when it emphasizes the omnipresence of symmetry, but it nevertheless makes a lively and fascinating read for a broad audience. ■

István Hargittai is a co-author (with Magdolna Hargittai) of *Symmetry Through the Eyes of a Chemist*. He is at the Budapest University of Technology and Economics, 1521 Budapest, Hungary.

BETTMANN/CORBIS

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Hard hearted? In *I, Robot*, Will Smith finds there is more to the robots than meets the eye.

Robots have feelings too

Who Needs Emotions? The Brain Meets the Robot

edited by Jean-Marc Fellous & Michael A. Arbib

Oxford University Press: 2005. 399 pp. £36.50, \$59.95

Dylan Evans

As with so much else in the field of artificial intelligence, the idea that robots might one day have emotions first appeared in science fiction. The Czech playwright Karel Capek, who coined the word 'robot' in 1921, pictured robots rebelling against their creators. Isaac Asimov went on to imagine robots with more positive emotions. Philosophers soon got in on the act and within a decade had explored many of the conceptual puzzles posed by these stories. With one or two notable exceptions, however, it wasn't until the 1990s that scientists and engineers finally began the attempt to turn the stories into reality.

Broadly speaking, research in emotional robotics can be divided into two distinct approaches. Some researchers prefer to concentrate on the practical task of giving robots the ability to interact with humans in emotional ways, such as detecting emotional states in people, or behaving in ways that are readily interpreted by people as expressions of emotion. Others set themselves the more ambitious task of endowing robots with an artificial analogue of the emotional-motivational system common to humans and many other animals. The two approaches reflect different goals: the first aims simply to produce robots that can interact socially with humans, whereas the second aims to deepen our understanding of what emotions really are.

Who Needs Emotions? deals mainly with the second approach. The book is a collection of

essays by some of the leading lights in affective neuroscience and emotional robotics. The editors, Jean-Marc Fellous and Michael Arbib, begin the book with a short dialogue between two imaginary characters, one with a theoretical bent, the other with a resolutely practical approach. This device allows them to introduce some of the key difficulties that beset the field in a remarkably concise and balanced manner. In just five pages, they give a glimpse of the thorny problems surrounding the definition of emotional terms and the difficulty of knowing when such terms apply to robots.

The two imaginary characters also serve as a way of classifying the other contributions to the book. Most of them are theoretical, with only a few describing actual attempts to build emotional robots. In the latter category, Ronald Arkin reviews a variety of simple behaviour-based robots that provide well defined models for a number of simple drives such as hunger, fear and sex, and Cynthia Breazeal and Rodney Brooks describe Kismet, an emotionally expressive robot that they designed to interact with people in a lifelike way. These are the most interesting chapters, because the existence of a real system, a concrete bit of hardware, allows for detailed questions to be asked about the differences and similarities between artificial and natural emotions.

The theoretical chapters present a variety of different models of emotion that aim both to describe how emotions work in biological creatures such as humans, and to serve as a blueprint for implementing emotions in robots. But these models lack any hardware implementation, so they are not easy to evaluate. In fact it is often difficult to see how they might be implemented, sometimes because they could be implemented in many different ways, and sometimes because they are simply

confusing. In the H-CogAff model proposed by Aaron Sloman, Ron Chrisley and Matthias Scheutz, for example, there are so many arrows in the diagram that it seems that every subsystem is connected with every other one, which calls into question the value of their distinction between reactive mechanisms, deliberative reasoning and meta-management. Andrew Ortony, Donald Norman and William Revelle also outline a tripartite model in which emotions are considered at three different levels of information processing — the reactive, the routine and the reflective — but these are not easily mapped on to the categories set out by Sloman, Chrisley and Scheutz.

This lack of theoretical agreement is characteristic of the whole field. Every researcher in this area has a pet theory, and yet there seems to be no agreement even as to the criteria that would allow us to choose one theory over another, let alone as to which theory is best. If some degree of consensus is a sign of scientific maturity, emotional robotics is still very much in its infancy.

But this is precisely what makes it such a difficult and interesting area of study. If you enjoy the kind of research that the philosopher of science Thomas Kuhn called "normal science", in which the ground rules are long established and it is now simply a question of solving well defined problems, this volume is not for you. But if you prefer the more disorienting and less predictable task of discovering the ground rules themselves, these essays on emotional robotics provide a sometimes frustrating, but always fascinating, glimpse of a science in the making. ■

Dylan Evans is a senior lecturer in intelligent autonomous systems at the University of the West of England, Frenchay Campus, Coldharbour Lane, Bristol BS16 1QY, UK.

ASTROPHYSICS

How to make a massive star

Barbara A. Whitney

Two competing theories have been applied to the formation of high-mass stars. Observations of two stellar systems now suggest that the accretion model has a weightier claim than its rival merger model.

How do high-mass stars form? Through accretion — the gravitational collapse of a dense cloud of gas and dust¹ — as do better-understood stars of lower mass like the Sun? Or do they form through the collision or merging of smaller stars? Two papers^{2,3} in this issue find evidence for dense disks of dust and gas around high-mass stars, adding to the mounting evidence for the first of the two options.

The origin of stars of high mass (ten or more times that of the Sun) is one of the most interesting unsolved questions in astrophysics. It has long been considered difficult, if not impossible, for these stars to form through accretion like their smaller cousins, because the intense radiation pressure from their central source halts the inflow of material. This, and the fact that young high-mass stars are often observed in very dense clusters of stars, led to the idea that they could have formed through the collision or merging of mid-sized stars⁴.

Both the accretion and merger models make predictions that can be tested observationally. In the accretion model, the effects of angular momentum in a star's progenitor cloud cause the formation of a disk and highly aligned, or 'collimated', supersonic outflows that clear out matter to form tunnels of lower density (Fig. 1a). According to the collision model, mergers are more likely to disrupt disks than to create them, and outflows are not likely to be collimated⁵ (Fig. 1b). The presence or absence of these features can thus be used to distinguish between the two models.

Patel *et al.* (page 109 of this issue)² and Jiang *et al.* (page 112)³ take very different approaches to find disks around two different massive stellar objects. In both cases, the disk in question is concealed inside a larger, less dense envelope of matter that obscures it from view at many typically observed wavelengths in the visible and near- to mid-infrared regions. The opacity of this dusty envelope decreases with increasing wavelength; so an obvious way round the problem is to observe at longer wavelengths.

This is the approach taken by Patel *et al.*² They observed an object called Cepheus A — with a mass about 15 times that of the Sun and

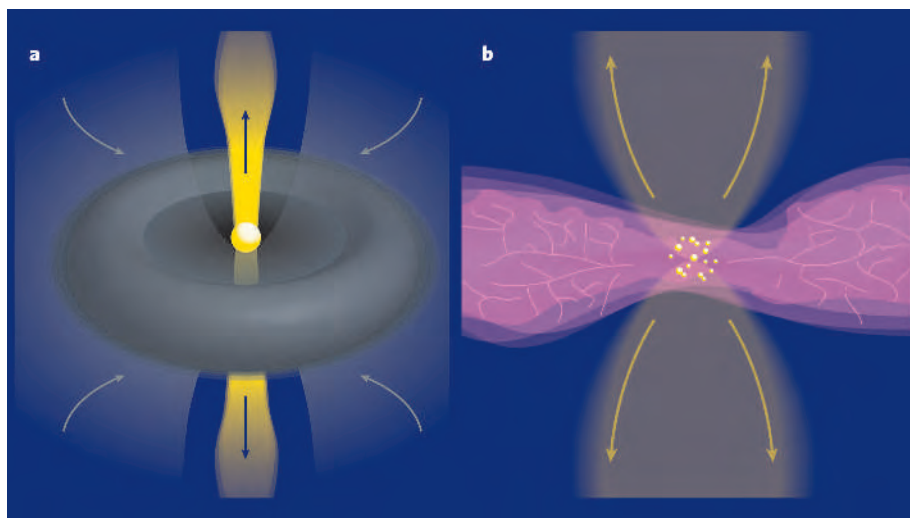


Figure 1 | Rival formations. **a**, A dense, rotating central disk of dust and gas, and a highly collimated supersonic outflow, are characteristic of the accretion model of high-mass star formation. An infalling envelope of gas also rotates, forming a torus around the disk. Outflows carve cavities (tunnels) of less dense material in this envelope. **b**, In the alternative merger model, many stars (including binary systems) are contained in the middle cloud centre, some of which collide and merge to form high-mass stars. Individual outflows and stellar winds from massive stars combine to give a large, uncollimated flow out from the system. Scales of **a** and **b** are about 1.5 light years across and 15 light years across, respectively.

a luminosity some 12,000 times greater — using the recently commissioned Submillimeter Array (SMA)⁶ telescope on Hawaii. The SMA allows unprecedented resolution at wavelengths sensitive to the dust and molecular emission from dense regions such as the disk around a star. Patel and colleagues observed in the far infrared at a wavelength of 917 μm , and found a disk with a radius of 300 AU (1 AU is the distance from the Sun to Earth), and a mass somewhere between one and eight times that of the Sun. The authors scanned across the surface of the disk, measuring changes in a characteristic wavelength of radiation emitted by the gas methyl cyanide, and concluded that the disk is rotating. The basis for this conclusion is a cosmic consequence of the well-known Doppler principle: things moving away emit sound or light waves at lower frequencies, and vice versa.

Observations of Cepheus A by the Very Large Array (VLA) telescope in New Mexico² at a longer radio wavelength of 3.6 cm supply further compelling evidence for accretion.

They show a highly collimated outflow in a direction perpendicular to the disk's plane, energized by hot, free electrons. Such mechanisms are ubiquitous in low-mass star formation, where to conserve angular momentum and allow most of the infalling gas and dust to accrete onto the star, some material is ejected out towards the polar regions.

The complementary observations of Jiang *et al.*³, made at a wavelength of 2 μm , concern the Becklin–Neugebauer (BN) object. At the centre of this object, familiar as one of the first big finds of infrared astronomy and as the brightest object in the sky (apart from the Sun) when observed at near-infrared wavelengths, is a star with seven times the mass of the Sun and 2,500 times its luminosity. The authors used a technique known as polarimetry to measure the scattering and polarization of light emitted by the central star. Because dust absorbs a large fraction of incident radiation at each point of scattering, denser matter in a region will absorb more light, making a polarized image appear darker. The image of the BN object made by

Jiang and colleagues shows a dark lane suggestive of a disk, accompanied by a brightening in two lobes, indicative of reduced scattering in less dense bipolar cavities.

Detailed models of radiative transfer can extract the geometry of the material surrounding the star from such pictures. These models include the effects not just of the scattering properties of dust, but also of the alignment of dust grains in magnetic fields, which impart polarization to radiation merely passing through and not scattering. Radiative transfer models of the BN source find a disk of radius 800 AU embedded inside an envelope roughly twice that size.

Because massive stars are typically more distant than low-mass stars, most previous observations of disk-like structures around massive stars did not have the resolution to distinguish objects of small angular extent. These observations most probably included the larger-scale 'envelope tori' surrounding the disk, or were taken at radio wavelengths that are sensitive to emission from outflows as well. They were thus not as convincing evidence for disks, and failed to clinch the argument for the accretion model.

Do the observations of Patel *et al.*² and Jiang *et al.*³ prove that the accretion model is correct? Perhaps — for the objects observed, at least. The observations provide a consistent picture of accretion that includes not just disks, but also rotating, infalling envelopes and outflows. These are just the ingredients that would, according to recent theories of accretion, allow the formation of massive stars: material that accretes into the cooler equatorial regions can be shielded from the extreme radiation of the star that normally halts the flow, whereas intense stellar radiation is channelled into the outflow cavities^{7,8}.

It may also be that both models of star formation operate, but in different circumstances: accretion in more isolated systems; and accretion, with merging, in dense clusters. Recent dynamical simulations indicate that the formation of close binary stars is a likely outcome of accretion in a cluster⁹. If the binaries are as close as 1 AU in separation, it is likely that some will merge. ■

Barbara A. Whitney is at the Space Science Institute, 4750 Walnut Street, Suite 205, Boulder, Colorado 80301, USA.
e-mail: bwhitney@spacescience.org

- Shu, F. H., Adams, F. C. & Lizano, S. *Annu. Rev. Astron. Astrophys.* **25**, 23–81 (1987).
- Patel, N. A. *et al. Nature* **437**, 109–111 (2005).
- Jiang, Z. *et al. Nature* **437**, 112–115 (2005).
- Bonnell, I. A. & Bate, M. R. *Mon. Not. R. Astron. Soc.* **336**, 659–669 (2002).
- Bally, J. & Zinnecker, H. *Astron. J.* **129**, 2281–2293 (2005).
- Ho, P. T. P., Moran, J. M. & Lo, K. Y. *Astrophys. J.* **616**, L1–L6 (2004).
- Yorke, H. W. & Sonnhalter, C. *Astrophys. J.* **569**, 846–862 (2002).
- Krumholz, M. R., McKee, C. F. & Klein, R. I. *Astrophys. J.* **618**, L33–L36 (2005).
- Bonnell, I. A. & Bate, M. R. *Mon. Not. R. Astron. Soc.* (in the press); preprint available at www.arxiv.org/astro-ph/0506689.

PALAEOLOGY

Between water and land

Robert L. Carroll

The most informative examples of large-scale evolution are provided by major transitions between environments. Fresh research on an ancient amphibian shows how it adapted to locomotion both in water and on land.

One of the defining events in the history of life was the emergence of terrestrial vertebrates from early fish. The oldest known fossils that illustrate the transition to land are those of *Ichthyostega* from the Upper Devonian of East Greenland. This 360-million-year-old amphibian resembled fish in many features of its skeleton, but possessed pelvic (hip) and pectoral (shoulder) girdles and limbs capable of supporting the body and allowing movement on land (Fig. 1). *Ichthyostega* was initially described by Säve-Söderbergh¹ and Jarvik^{2,3}, but many aspects of its anatomy remained unknown, limiting understanding of aspects of the transition from water to land.

On page 137 of this issue, Ahlberg, Clack and Blom⁴ provide a new reconstruction and functional analysis based on the original specimens, as well as on many fossils collected subsequently by Clack and her colleagues. Their work reveals numerous specializations in the components of the main body axis — the axial skeleton — that were not previously recognized. They demonstrate that *Ichthyostega* possessed a unique vertebral structure that enabled the trunk to be supported above the ground, and that it also had an elaboration of the ribs at the base of the tail that facilitated swimming.

Unlike other primitive amphibians and their plausible antecedents among lobe-finned fish, the configuration of the trunk vertebrae in *Ichthyostega* is not uniform along the column, but instead is regionally specialized. Much of the interpretation hinges on variation in the vertebrae, each of which consists of a crescentic ventral portion (the centrum) partially surrounding the notochord, and a neural arch that encloses the spinal cord and

extends dorsally as a neural spine.

In *Ichthyostega*, the neural spines of the vertebrae behind the shoulder girdle are angled backwards, while those above and in front of the pelvic girdle are angled forwards. Those in the middle of the trunk are more nearly vertical, but have pronounced areas of roughness, as occur in modern vertebrae for the attachment of muscles and associated tendons. Because muscles produce their maximum force when acting at right angles to the structures to which they are attached, the varied geometry of the neural spines indicates that the attached muscles could have raised the central portion of the trunk, lifting it above the ground. This configuration is analogous to that of a suspension bridge, with the trunk lifted by muscles (acting like cables) running from elevated supports (the pelvic and pectoral girdles and limbs).

Ahlberg *et al.*⁴ point out that the widely expanded anterior ribs, also unique to *Ichthyostega*, would have made the spinal column more rigid, but would also have greatly restricted lateral undulation of the trunk, which is a major factor in aquatic locomotion. The great length of these ribs, compared with their size in lobe-finned fish, would also have prevented collapse of the chest cavity and lungs as *Ichthyostega* dragged itself out of the water. The rigidity of the anterior trunk appears to have been compensated for by enlargement of the ribs at the base of the tail, which would have increased the area for attachment of muscles that moved the tail from side to side.

The lineage that included *Ichthyostega* thus seems to have undergone the transition from water to land through changes that would have

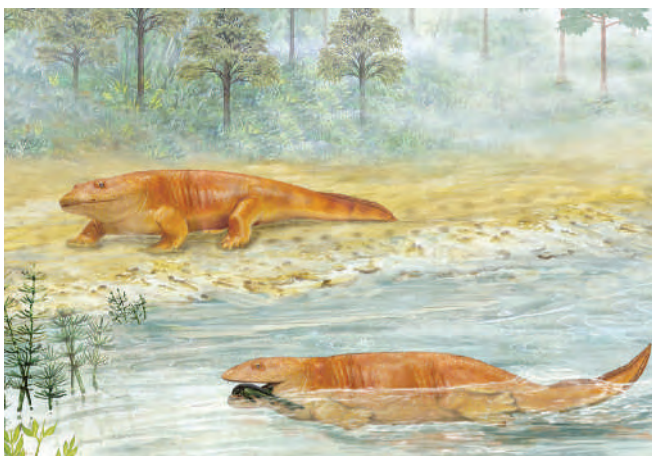


Figure 1 | Reconstructions of the body form of *Ichthyostega*, one of the oldest known amphibians. Length of the animals is about 1 m. These depictions show their capacity for swimming in water using the tail, or walking on land with the trunk supported by muscles attached to specialized neural spines on the vertebrae. (Reconstruction by Tony Terenzi, based on ref. 4.)

increased the facility for movement in both environments. What remains to be learned of the skeleton of *Ichthyostega* includes the anatomy of the wrist and hand, and the orientation of the rear limb when walking on land.

Surprisingly, the particular design observed in *Ichthyostega* was ultimately unsuccessful, for there are few if any fossils known after about 360 million years ago that represent plausible descendants of this lineage. Rather than elaboration of the neural spines and their musculature to support the vertebral column, the more successful terrestrial vertebrates expanded the vertebral centra, the same elements that provide the major support in all modern terrestrial vertebrates. Homologues of the centra in human vertebrae were little, if at all, ossified in *Ichthyostega*. It is interesting to note that the development of the different elements of the vertebrae is under different genetic control. The dorsal and lateral portions of the neural arch are controlled by bone morphogenetic protein 4 (BMP-4) and the vertebral centra by Pax1 (ref. 5). Random mutations in one or the other of these developmental pathways could have led to alternative directions of evolutionary change.

From the standpoint of patterns of evolution, there are at least 11 lineages of advanced lobe-finned fish and early amphibians from the Upper Devonian⁶, only one of which is a plausible close relative of later land vertebrates. What one sees from the fossil record is a 15-million-year history of 'experimentation' among the descendants of fish that had already evolved fins with a central bony axis, a swim bladder (the fish equivalent of lungs) and paired nostrils opening into the mouth⁷. Comparable diversity of ultimately unsuccessful experiments is also evident in other major transformations in the evolution of vertebrates, including the early evolution of mammals⁸, the origin of whales⁹ and evolution within our own genus *Homo*¹⁰. In the last case, as many as ten recognized species have diverged within the past 2 million years, differing greatly in stature, cranial capacity and facial structure, only one of which remains to give rise to future generations. ■

Robert L. Carroll is at the Redpath Museum, McGill University, 859 Sherbrooke Street West, Montreal, Quebec H3A 2K6, Canada.
e-mail: robert.carroll@mcgill.ca

PALAEOCLIMATE

The riddle of the sediments

Mark Siddall

The ratio of oxygen isotopes contained in the signal in deep-sea sediments can tell us a great deal about past ice-volume variations. The challenge is to disentangle the different contributions to the signal.

The ratio of ice to water on Earth has been in more-or-less constant flux on timescales of thousands to millions of years. Although we have known of these changes for several decades, quantifying them — finding whether Earth's climate has laid down a precise account of itself anywhere — continues to preoccupy palaeoceanographers, palaeoclimatologists and glaciologists alike. Richard Bintanja and colleagues (page 125 of this issue)¹ have used a novel and elegant method to investigate further one possible source for such a record: the ratio of oxygen isotopes contained in benthic, or deep-sea, foraminifera.

Benthic foraminifera are the fossilized remains of single-celled organisms that lived on or close to the ocean floor. The ratio of the oxygen isotopes ¹⁸O and ¹⁶O in the shells of benthic foraminifera varies with the depth of the sediments in which they are buried. The fact that this variation is similar at sites throughout the world encourages the suspicion that it originates from climate-related changes in Earth's oxygen-isotope budget. As temperatures decrease and glaciers form, proportionately more of the lighter oxygen isotope ¹⁶O tends to become incorporated into ice. The ratio ¹⁸O/¹⁶O in the remaining sea water thus climbs in an approximately linear fashion with respect to increases in glacier volume².

This idea is simple enough. But there are complications when it comes to extracting the effect of ice-cycles on the isotopic ratio from the information contained in the shells of benthic foraminifera. First, the relative degree of inclusion of the two oxygen isotopes into the shells varies according to the deep-ocean temperature where and when the shell formed. Second, differences in the isotopic ratio between water masses, coupled with changes in deep-ocean circulation, in turn affect the ratio of ¹⁸O to ¹⁶O incorporated into the shells of benthic foraminifera. Previous attempts to disentangle these effects and separate out the ice-cycle signal in oxygen-isotope ratios of benthic foraminifera have had some limited success^{3–5}. But until now, no work has actually verified whether the ice volumes suggested by these ratios are compatible with estimates of temperature at the sites of the ancient, continental ice sheets and the dynamics of ice-sheet growth.

Bintanja and colleagues' work¹ changes that. Their central assumption is that deep-ocean temperatures are related to the surface-air temperatures at high latitudes in the Northern

Hemisphere by a simple, linear transformation. This enables them to link algorithms that describe temperatures in the deep ocean to those that calculate the ratios of oxygen isotopes in the ice sheets of the Northern Hemisphere. The authors then adjust the linked benthic and high-latitude temperatures of their coupled model, iteratively, to generate the right amount of ice growth and benthic temperature to best fit the record of benthic oxygen isotope ratios, and thus arrive at mutually consistent records of temperature and ice volume at high latitudes in the Northern Hemisphere.

To link deep-ocean and high-latitude surface-air temperatures seems quite a leap of faith. There is, however, some justification for the assumption. The waters of the deep ocean originate at high latitudes and sink by convection; and convection is, to some extent, a temperature-dependent process. The simplicity of Bintanja and colleagues' link is also quite possibly its strength: if a simple parametrization misbehaves, it is more likely to present obvious discrepancies in any exercise designed to validate it. In this case, comparison with independent sea-level reconstructions and temperature records from Antarctic ice cores (the only ice records long enough for such a comparison) reveals no such discrepancies. A further verification, not shown explicitly in Bintanja and colleagues' paper, is their prediction of the temperature variation in the deep ocean between glacial and interglacial periods: at 2–3 °C, this is in good agreement with independent estimates^{3–7}. (This range can be calculated simply by taking that part of the variation in oxygen isotopic ratio, in parts per thousand, that can be attributed to deep-sea temperature fluctuations, and multiplying it by a factor of 0.21 °C.)

Another challenge for Bintanja *et al.*¹ was to choose the benthic isotope record they should use to drive their model, as not all such records contain the same temperature or oxygen isotope signal^{5,8}. Their solution was to compile results from different ocean basins and latitudes, assuming that this process would average out local effects. An alternative approach is to choose sampling sites carefully so as to either maximize or minimize local variation. The deep Pacific is, for example, thought to be the ocean with the smallest local temperature effects. But if high-latitude surface-air and deep-ocean temperatures do indeed scale approximately linearly, Atlantic records of oxygen isotope variation may be the most

1. Säve-Söderbergh, G. *Medd. Grönland* **98**, 1–211 (1932).
2. Jarvik, E. *Basic Structure and Evolution of Vertebrates* (Academic, London, 1980).
3. Jarvik, E. *Fossils Strata* **40**, 1–206 (1996).
4. Ahlberg, P. E., Clack, J. A. & Blom, H. *Nature* **437**, 137–140 (2005).
5. Christ, B., Huang, R. & Scaal, M. *Anat. Embryol.* **208**, 333–351 (2004).
6. Clack, J. *Amphibian Biology* Vol. 4. 979–1029 (Surrey Beatty, Chipping Norton, NSW, Australia, 2000).
7. Zhu, M. & Ahlberg, P. *Nature* **432**, 94–96 (2004).
8. Luo, Z.-X., Kielan-Jaworowska, Z. & Cifelli, R. *Acta Palaeontol. Pol.* **47**, 1–78 (2002).
9. Thewissen, J. & Williams, E. M. *Annu. Rev. Ecol. Syst.* **33**, 73–90 (2002).
10. Lahr, M. & Foley, R. *Nature* **431**, 1043–1044 (2004).

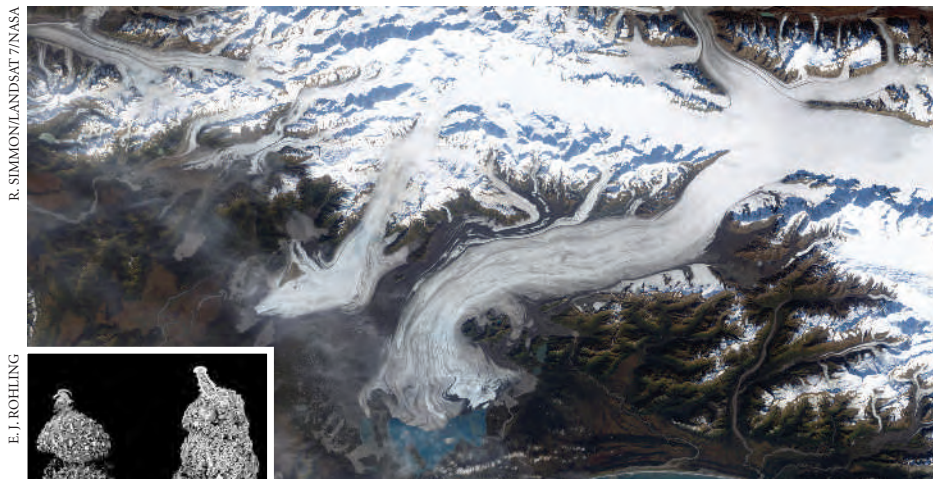


Figure 1 | Simple link? Bintanja and colleagues' approach¹ connects deep-ocean temperatures, as inferred from oxygen-isotope ratios in the remains of benthic foraminifera (inset), with temperatures at high latitude in the Northern Hemisphere and glacier formation.

suitable — after all, convection in the Atlantic may be the link between high-latitude surface-air temperatures and deep-sea temperatures. Bintanja *et al.* tackle the issue of the correct choice of benthic record in the Supplementary Information to their paper. Some differences between model runs with different benthic isotope records are apparent, but in general their method is robust for the alternative cores they use.

This paper fires the imagination with ways to test the ideas contained in it and, if those ideas are right, directions in which to extend the work. The central assumption — of a simple link between deep-ocean temperature and high-latitude Northern Hemisphere temperature (Fig. 1) — is a well defined hypothesis that should be eminently testable using Earth System Models of Intermediate Complexity (EMICS). But how far into the past can we push this method to understand Earth's glacial cycles during different climatic periods? As benthic oxygen isotope records of higher

temporal resolution become available, what will this approach have to say about periods of millennial sea-level variability?

Bintanja and co-workers' approach¹ provides an elegant new way of deciphering the riddle of ice volume change contained in the benthic record. It is now up to the various interested communities to take up the challenges posed by this work. ■

Mark Siddall is in the Department of Climate and Environmental Physics, University of Bern, Sidlerstrasse 5, 3012 Bern, Switzerland.
e-mail: siddall@climate.unibe.ch

1. Bintanja, R., van de Wal, R. S. W. & Oerlemans, J. *Nature* **437**, 125–128 (2005).
2. Shackleton, N. J. & Opdyke, N. D. *Quat. Res.* **3**, 39–55 (1973).
3. Chappell, J. & Shackleton, N. J. *Nature* **324**, 137–140 (1986).
4. Shackleton, N. J. *Quat. Sci. Rev.* **6**, 183–190 (1987).
5. Waelbroeck, C. *et al.* *Quat. Sci. Rev.* **21**, 295–305 (2002).
6. Adkins, J. F. *et al.* *Science* **298**, 1769–1773 (2002).
7. Cutler, K. B. *et al.* *Earth Planet. Sci. Lett.* **206**, 253–271 (2003).
8. Labeyrie, L. D. *et al.* *Nature* **327**, 477–482 (1987).

CANCER IMMUNOLOGY

Cat and mouse games

Cornelis J. M. Melief

The immune system is intimately involved in how tumours develop. But how do tumours avoid being killed by immune responses? It seems that in some instances they can lull immune cells into a false sense of security.

A troubled relationship exists between the immune system and tumours — not unlike that between the cartoon characters Tom and Jerry. Tom's disposition is to catch and devour mice, but he hardly ever succeeds in capturing Jerry, who has just too many tricks up his sleeve. On page 141 of this issue Willimsky and Blankenstein¹ report a novel ploy used by

tumours to evade the immune system. It seems that sporadic tumours in mice, against which the immune system initially reacts, nevertheless manage to spread by moulding the immune cells that are most effective against them — the killer T cells — into a state that tolerates them.

The report adds fuel to a controversy that has been raging for some time, namely



50 YEARS AGO

Wittgenstein once argued that no paper read to a scientific society should last more than fifteen minutes, because anyone who had anything to say on any subject could easily say it in that time. He would probably have regarded with some horror the Twelfth International Congress of Applied Psychology, which met in London during July 18–23. Few of the speakers named on the programme took less than half an hour, and there were more than a hundred of them. Yet only four special lecturers had been asked to talk for more than twenty minutes. Perhaps many of the others felt that, having come from the ends of the earth, or thereabouts, they could justify themselves... only by going on and on. Nor was the unbargained-for time always well spent. In fact, and as usual, the people who spoke the longest had the least to say.

From *Nature* 3 September 1955.

100 YEARS AGO

Mr. W. E. Cooke, Government astronomer for Western Australia, has sent us a communication explaining a novel plan that he has adopted for giving more definiteness to the weather forecasts issued in that colony. Each forecast for a definite district is subdivided into specific items, to each of which a figure is attached, "1" representing that the occurrence prognosticated has only the barest possibility of being successful, and so on, up to "5", which indicates that the prediction may be relied upon with almost absolute certainty. Each item of the forecast has therefore a "weight" attached to it; on the whole, Mr. Cooke states that the new method has proved a distinct success, and that while people find that whenever the figure 5 appears the forecast is fulfilled in 99 cases out of 100, they do not feel so disappointed in case of failure when the lower numbers are attached...

From *Nature* 31 August 1905.

50 & 100 YEARS AGO

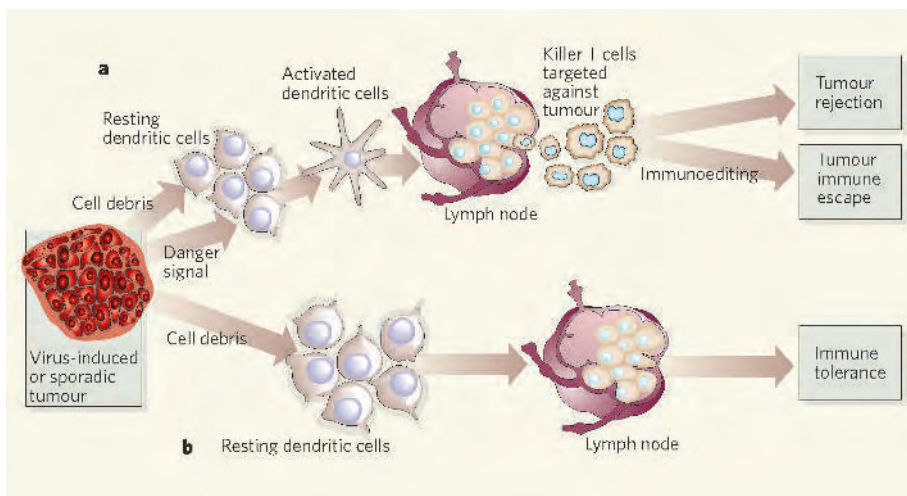


Figure 1 | Escaping the immune system — a model. After initial growth, tumours usually shed some immunogenic material from dead or dying tumour cells. This debris is picked up by dendritic cells, which transport it to the lymph node and 'present' it to T cells. The subsequent immunological events are determined by the manner in which the tumour is perceived by the dendritic cell network. **a**, If the tumour, apart from shedding debris, also emits 'danger' signals such as stress proteins, the dendritic cells will be activated. These activated cells present the tumour debris to the T cells, eliciting a robust response and causing the T cells to multiply and kill tumour cells. The only way for tumour cells to survive is to escape by immunoediting^{2,7}. **b**, If the tumour manages to masquerade as healthy tissue, giving off no danger signals, the dendritic cells are not activated. The T cells therefore tolerate the tumour material presented to them, and do not become killers⁹. Tumours capable of such tolerance induction do not need immunoediting to escape from immune attack. Tumours that are induced by viral infection are more likely to fall into the first category; Willimsky and Blankenstein's¹ mouse model seems to produce tumours that fall into the second.

whether the development of sporadic, non-viral tumours is really affected by so-called immunoediting. The theory of immunoediting posits that tumours eliciting strong immune responses from the T cells of their host will experience a darwinian selection pressure to avoid destruction by the immune system. So, they evolve 'immune escape' or 'sculpted' variants, usually by losing or down-regulating the molecules that alerted the immune system in the first place^{2,3}.

In Willimsky and Blankenstein's experiment, however, it is not the tumour cells that adapt to immune attack — rather, the immune system is coaxed into inertia. The authors genetically engineered an ingenious mouse model, which contained a cancer-promoting gene of viral origin (encoding the 'SV40 large T' protein). The gene was controlled so that it was activated only rarely and in random tissues. The mice thus developed rare sporadic tumours, all expressing SV40 large T. Although this protein initially provoked an immune response in the mice, in the course of tumour development the mice became immunologically tolerant of it. However, the tumours were vigorously rejected when transplanted into identical mice without tumours — so they could still elicit a tumour-killing immune response¹ — and they showed no obvious signs of immunoediting^{1,2}.

Curiously, both sides of the controversy concur that both virus-induced and sporadic tumours do elicit immune responses, and that T-cell responses protect against viral carcinogenesis. An example of the latter is the

markedly increased incidence of cancers induced by Epstein–Barr virus in AIDS patients and in immunosuppressed recipients of organ transplants⁴. These tumours also express a much wider array of viral antigens⁴. But what distinguishes the immune response against virus-induced cancers from that against sporadic tumours? After all, Willimsky and Blankenstein's model was based on sporadic carcinogenesis, but using the viral SV40 large T protein. This protein is one of the main targets of successful T-cell responses against tumours induced by intact SV40, the virus from which the large T protein is derived. Indeed, the efficient T-cell reactions against this antigen are instrumental in the prevention of tumour development in mice that have been injected with intact SV40 (ref. 5).

Willimsky and Blankenstein offer little explanation for the striking difference in the protective power of immune responses against the identical viral antigen. And they conclude from their results that cancers that are not induced by viruses do not undergo immunoediting. But they fail to take into account the extensive documentation from others that such tumours are in fact subject to selective pressure from immune responses^{2,3}.

How can these seemingly contradictory facts and widely divergent views be reconciled? Two points are worth emphasizing here. First, tumours need to be sensed by an exquisitely sensitive sentinel system, the dendritic cell network (Fig. 1). These cells not only sense the tumours, they also direct the subsequent immune response. Dendritic cells pick up

material from dead or dying tumour cells, and 'present' it to T cells. For proper arousal of killer T cells, the dendritic cells must be activated by secondary 'danger' signals, some of which can be emitted by tumours. The presentation of the tumour debris by the activated dendritic cell causes T cells to multiply and attack the remaining tumour cells⁶. However, if the dendritic cells are not activated, the T cells will take little notice and can become tolerant of the tumour^{6–8}.

Second, the natural ability of tumours to activate dendritic cells is likely to vary widely, from no stimulatory activity to considerable stimulatory action through endogenous danger signals such as overexpressed stress proteins and interferons (proteins that were named for their capacity to interfere with viral infections but are also produced in response to non-viral tumours)^{6,9}. So an important distinction should be made between tumours whose growth is associated with sufficient dendritic-cell activation to generate killer T cells targeted against them, and tumours that do not (Fig. 1). Immunoediting should occur only in the first category, because escape is needed only in the face of immune attack.

Although virus-associated tumours are more likely to cause dendritic-cell activation than sporadic tumours, and are therefore more likely to be subjected to immunoediting, this model seems to be too simple. Many sporadic, non-viral tumours can apparently arouse enough of the body's danger signals to become subject to immune attack and undergo immunoediting^{2,3,9}. Conversely, most tumour viruses fail to incite sufficient danger signals in susceptible individuals and consequently can establish the persistent viral infection associated with a high risk of cancer¹⁰.

So, the bad news is that cancer cells can avoid robust T-cell-mediated immune attack, either by failing to arouse the immune system, thereby causing tolerance¹, or by immunoediting^{2,9}. The good news is that T cells can be turned into robust killers by the vigorous activation of dendritic cells using a variety of molecules⁸, and this may be of use in cancer vaccines^{6,8} and in T-cell immunotherapy¹¹. ■

Cornelis J. M. Melief is in the Department of Immunohematology and Blood Transfusion, Leiden University Medical Center, Albinusdreef 2, PO Box 9600, 2300 RC Leiden, The Netherlands.

e-mail: c.melief@lumc.nl

1. Willimsky, G. & Blankenstein, T. *Nature* **437**, 141–146 (2005).
2. Dunn, G. P. et al. *Nature Immunol.* **3**, 991–998 (2002).
3. Khong, H. T. & Restifo, N. P. *Nature Immunol.* **3**, 999–1005 (2002).
4. Straathof, K. C. M. et al. *Oncologist* **8**, 83–98 (2003).
5. Newmaster, R. S., Mylin, L. M., Fu, T.-M. & Tevethia, S. S. *Virology* **244**, 427–441 (1998).
6. Melief, C. J. M. *Nature Immunol.* **6**, 543–544 (2005).
7. Probst, H. C. et al. *Nature Immunol.* **6**, 280–286 (2005).
8. Dumortier, H. et al. *J. Immunol.* **175**, 855–863 (2005).
9. Dunn, G. P. et al. *Nature Immunol.* **6**, 722–729 (2005).
10. De Jong, A. et al. *Cancer Res.* **64**, 5449–5455 (2004).
11. Klebanoff, C. A. et al. *Proc. Natl Acad. Sci. USA* **102**, 9571–9576 (2005).

OBITUARY

John Norris Bahcall 1935–2005

Nuclear astrophysicist who uncovered the solar neutrino problem.

The philosopher Sir Isaiah Berlin once famously quoted a scrap of Archilochus: “The fox knows many things, but the hedgehog knows one big thing.” Astrophysicist John Bahcall would often introduce himself and a colleague to a new acquaintance with the sentence: “I know all about neutrinos, and my friend here knows about everything else in astrophysics.”

Such light-hearted self-deprecation was typical of Bahcall, who died on 17 August 2005. But it was inaccurate: Bahcall’s scientific interests and expertise ranged from neutrino physics and the structure of the Sun and other stars, to galaxy models, quasars and the intergalactic medium. His more than 600 scientific publications, on an enormous array of subjects, received nearly 20,000 citations. Many established fundamental paradigms in their fields, and others provided the clearest and most comprehensive overview of them.

But Bahcall did, in fact, continually return to one core scientific issue: the solar neutrino problem. He realized very early in his career that we should be able to detect the flux, or stream, of these shadowy fundamental particles as they pass through the Earth after escaping from the centre of the Sun, where they are produced in prodigious numbers. He clearly saw that a definite detection, or non-detection, of these neutrinos would have major implications both for understanding the Sun and for fundamental particle physics. For decades, he encouraged and supported scientists throughout the world in studying this problem and was most successful in his collaboration with Raymond Davis Jr, who ultimately won the Nobel Prize in 2002 for detecting the solar neutrino flux.

It was Bahcall’s persistent work that proved definitively that the low flux found by the solar neutrino experiments of Davis and others could not be explained by errors in our model for the Sun. Neutrinos seemed to be missing: either they were not made at the rates required by standard nuclear physics, or they were made but then somehow ‘lost’ in transit between the Sun and the Earth. The latter explanation — neutrino mixing, in which one type of neutrino changes into another at some rate, and in which the neutrino must have a small but finite mass — is now known to be true, and it is surely due to Bahcall’s tenacity and insight that this important and surprising modification to the standard model of particle physics was uncovered.

A fuller idea of his exceptional scientific scope is indicated by the fact that the standard model for a massive black hole surrounded

by a cluster of stars is still called the Bahcall–Wolf model; the most widely quoted model for our Galaxy was for decades the Bahcall–Soneira model; and the most accurate models for the solar interior were those developed by Bahcall, with Roger Ulrich, Marc Pinsonneault and others.

After completing his graduate education at Harvard in 1961 (having started his university education at Louisiana State University on a tennis scholarship), he arrived at the California Institute of Technology (Caltech), working with Willy Fowler and others at the time and place that ‘nuclear astrophysics’ was invented. There he became engaged with neutrino work and to Neta Assaf (then completing her PhD at Caltech) — the two constant loves of his life. His first paper from Caltech, a one-page letter to the editor of the *Astrophysical Journal*, dated 1 December 1962 and entitled “The Solar Neutrino Flux” (authored with Fowler, Icko Iben and Richard Sears), proposed an experiment that might “provide a valuable experimental limit on the effective temperature for neutrino generation in the sun”. That paper set the course for a lifetime of research.

The writing of scientific papers was, however, only one of Bahcall’s many contributions to world science. He was an educator who changed the nature of postdoctoral training, and a scientific statesman of unusual and beneficent influence. Bahcall moved to the Institute for Advanced Study (IAS) at Princeton in 1968 and soon established that institution as a magnet and model for postdoctoral training. A significant fraction of the world’s most distinguished astrophysicists benefited from his tutelage and the intellectually fertile atmosphere that he established there. The eminent British scientist Sir Martin Rees describes himself as fortunate to have been one of the first IAS postdoctoral fellows in astrophysics in 1969. The birthdays of all the fellows and important family events were celebrated. The intellectual atmosphere was intense, and the weekly Tuesday lunches, with John presiding, to which the whole Princeton physics community was invited, were legendary. Bahcall’s postdoc programme was the one that astrophysics institutions worldwide emulated. At the IAS, young scientists were selected and recruited in the most exacting manner and then were free to work on whatever they wanted, with whomever they wished.

While maintaining a scientific and educational programme that would have exhausted most, Bahcall also demonstrated extraordinary scientific leadership. He was president of the American Astronomical Association, led



C. MOORE

the team that produced the 1990 National Research Council ‘Bahcall Report’ that set the scientific and instrumental priorities for astrophysics in the United States for a decade, and worked (with Lyman Spitzer Jr) with tireless effectiveness in public and in private to have the Hubble Space Telescope and the Space Telescope Science Institute (STScI) built and maintained as one of the world’s pre-eminent scientific facilities.

Neta Bahcall, a distinguished scientist herself, was his love, his best friend and his scientific colleague throughout. She took a leading scientific role at the STScI and wrote over 30 papers with him on subjects ranging from solar neutrinos to binary X-ray sources. They also collaborated in raising three talented children, Safi, Dan and Orli, who are themselves now establishing significant scientific careers.

But no listing of achievements can convey the impression of the man: the wit, the mischievous energy, the passion. Jerry Wasserburg, his old Caltech friend, portrays Bahcall in 1965: “John, running around in white tennis shorts, very sportive and competitive in both creative science and tennis, trying out and enthusiastically arguing every new idea in astrophysics; he was the dynamo of the institute.”

Jeremiah P. Ostriker

Jeremiah P. Ostriker is in the Department of Astrophysical Sciences, Princeton University, Peyton Hall, Ivy Lane, Princeton, New Jersey 08544-1001, USA.
e-mail: ostriker@princeton.edu

BRIEF COMMUNICATIONS

Ibuprofen-like activity in extra-virgin olive oil

Enzymes in an inflammation pathway are inhibited by oleocanthal, a component of olive oil.

Newly pressed extra-virgin olive oil contains oleocanthal — a compound whose pungency induces a strong stinging sensation in the throat, not unlike that caused by solutions of the non-steroidal anti-inflammatory drug ibuprofen¹. We show here that this similar perception seems to be an indicator of a shared pharmacological activity, with oleocanthal acting as a natural anti-inflammatory compound that has a potency and profile strikingly similar to that of ibuprofen. Although structurally dissimilar, both these molecules inhibit the same cyclooxygenase enzymes in the prostaglandin-biosynthesis pathway.

The agent in extra-virgin olive oil responsible for throat irritation is thought to be the dialdehydic form of (–)deacetoxy-ligstroside aglycone² (or oleocanthal, with oleo- for olive, -canth- for sting, and -al for aldehyde) (Fig. 1). To confirm this, we isolated (–)oleocanthal from different premium olive oils and measured its intensity as a throat irritant. We found that irritation intensity was positively correlated with oleocanthal concentration. Although this finding indicates that oleocanthal is probably the principal irritating compound in olive oil, it was possible that

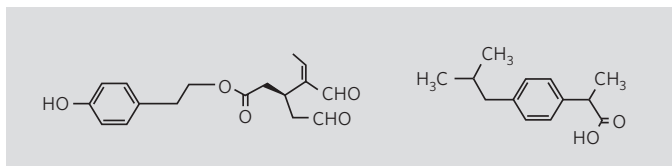


Figure 1 | Structures of (–)oleocanthal (left) and the anti-inflammatory drug ibuprofen (right). How they underpin the similar throat-irritating and pharmacological properties of the two compounds is unclear as yet.

co-elution of a minor component or a mixture of components could be causing the burning sensation². We therefore completed a *de novo* synthesis of oleocanthal, assigned the absolute stereochemistry (A.B.S. and Q.H., unpublished results), and tested the throat-irritant properties of this synthetic (–)oleocanthal, dissolved in non-irritating corn oil. The effect was comparable to that of premium extra-virgin olive-oil oleocanthal and was also dose-dependent. (For details and for methods, see supplementary information.)

Forty years ago, it was found that the bitterness of certain compounds correlated with their pharmacological activity³. On the basis of their shared irritant properties, we therefore tested whether oleocanthal might mimic the pharmacological effects of ibuprofen (Fig. 1), a potent modulator of inflammation and analgesia⁴. Ibuprofen is a non-selective

inhibitor of the cyclooxygenase enzymes COX-1 and COX-2, but not of lipoxygenase⁴, which catalyse steps in the biochemical inflammation pathways derived from arachidonic acid. We found that, like ibuprofen, both enantiomers of oleocanthal caused dose-dependent inhibition of COX-1 and COX-2 activities but

had no effect on lipoxygenase *in vitro* (Table 1).

Our findings raise the possibility that long-term consumption of oleocanthal may help to protect against some diseases by virtue of its ibuprofen-like COX-inhibiting activity^{5,6}. If 50 g of extra-virgin olive oil containing up to 200 µg per ml oleocanthal is ingested per day⁷, of which 60–90% is absorbed^{8,9}, then this corresponds to an intake of up to 9 mg per day. This dose is relatively low, corresponding to about 10% of the ibuprofen dosage recommended for adult pain relief, but it is known that regular low doses of aspirin, for example, another COX inhibitor, confer cardiovascular health benefits¹⁰. Ibuprofen is associated with a reduction in the risk of developing some cancers⁵ and of platelet aggregation in the blood¹¹, as well as with the COX-independent secretion of amyloid-β42 peptide in a mouse model of Alzheimer's disease⁶. A Mediterranean diet, which is rich in olive oil, is believed to confer various health benefits, some of which¹² seem to overlap with those attributed to non-steroidal anti-inflammatory drugs.

Our discovery of COX-inhibitory activity in a component of olive oil offers a possible mechanistic explanation for this link.

Gary K. Beauchamp*, **Russell S. J. Keast*†‡**, **Diane Morel‡**, **Jianming Lin§**, **Jana Pika§**, **Qiang Han||**, **Chi-Ho Lee*†**, **Amos B. Smith*||**, **Paul A. S. Breslin***

*Monell Chemical Senses Center, ‡Department of Pharmacology and Toxicology, University of the Sciences in Philadelphia, and ||Department of Chemistry, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA
e-mail: breslin@monell.org

†Present addresses: Food Science, RMIT University, Melbourne, Victoria 3001, Australia (R.S.J.K.); Animal Resources Research Center, Konkuk University, Gwangjin-Gu, Seoul 143-130, South Korea (C.-H.L.)
§Firmenich, PO Box 5880, Princeton, New Jersey 08543, USA

Table 1 | Selective inhibition of COX enzymes by oleocanthal enantiomers

Agent	Concentration (µM)	COX-1 (%)	COX-2 (%)	15-LO (%)
(–)Oleocanthal	100	83.5 ± 3.5	70.9 ± 8.6	0.4 ± 0.8
	25	56.1 ± 3.2	56.6 ± 9.5	0.0 ± 0.0
	7	24.6 ± 7.3	14.5 ± 2.3	0.0 ± 0.0
(+)Oleocanthal	100	68.0 ± 15.2	69.6 ± 3.9	3.5 ± 5.5
	25	54.5 ± 4.6	41.3 ± 15.9	0.7 ± 1.0
	7	24.6 ± 7.5	6.1 ± 4.2	0.0 ± 0.0
Ibuprofen	25	17.8 ± 2.3	12.7 ± 3.6	0.2 ± 0.3
	7	0.0	1.3	ND
Indomethacin	25	90.1	89.8	0.1 ± 0.9
	7	86.6	66.3	0.5 ± 0.1
NDGA	25	ND	ND	63.1 ± 0.8
	7	ND	ND	52.5 ± 1.1
caffeic acid	25	ND	ND	25.2 ± 2.2
	7	ND	ND	19.8 ± 1.3

Percentage inhibition of cyclooxygenases 1 and 2 (COX-1, COX-2) and 15-lipoxygenase (15-LO) by three different concentrations of oleocanthal and of ibuprofen are presented as mean ± s.e.m. for three independent experiments. ND, not determined. Indomethacin was used as a positive (inhibitory) control in the cyclooxygenase assays and *nor*-dihydroguaiaretic acid (NDGA) and caffeic acid were used as positive (inhibitory) controls in the lipoxygenase assay. The concentrations for 50% inhibition (IC₅₀; calculated by least-squares regression analysis of inhibition versus concentration) for the natural (–) oleocanthal are 23 µM and 28 µM for COX-1 and COX-2, respectively; IC₅₀ values for (+) oleocanthal are 25 µM and 40 µM for COX-1 and COX-2, respectively; IC₅₀ values for ibuprofen are 5 µM and 223 µM for COX-1 and COX-2, respectively¹³. (For methods, see supplementary information.)

- Breslin, P. A. S., Gingerich, T. N. & Green, B. G. *Chem. Sens.* **26**, 55–66 (2001).
- Andrewes, P., Busch, J. L. H. C., De Jooide, T., Groenewegen, A. & Alexandre, H. J. *Agric. Food Chem.* **51**, 1415–1420 (2003).
- Fischer, R., Griffen, F., Archer, R. C., Zinsmeister, S. C. & Jastram, P. S. *Nature* **207**, 1049–1053 (1965).
- Vane, J. R. & Botting, R. M. *Inflamm. Res.* **44**, 1–10 (1995).
- Harris, R. E., Beebe-Donk, J., Doss, H. & Burr Doss, D. *Oncol. Rep.* **13**, 559–583 (2005).
- Zhou, Y. et al. *Science* **302**, 1215–1217 (2003).
- Hu, F. N. *Engl. J. Med.* **348**, 2595–2596 (2003).
- Tuck, K., Freeman, M., Hayball, P., Stretch, G. & Stupans, I. *J. Nutr.* **131**, 1993–1996 (2001).
- Miro-Casas, E. et al. *Clin. Chem.* **49**, 945–952 (2003).
- Hennekens, C. H. *Am. J. Man. Care* **8** (suppl.), 691–700 (2002).
- Lefer, A. M., Muller, H. F. & Smith, J. B. *Br. J. Pharmacol.* **83**, 125–130 (1984).
- Togna, G. I., Togna, A. R., Franconi, M., Marra, C. & Guiso, M. J. *Nutr.* **133**, 2532–2536 (2003).
- Mitchell, J. A. et al. *Proc. Natl Acad. Sci. USA* **90**, 11693–11697 (1993).

Supplementary information accompanies this communication on Nature's website.
Competing financial interests: declared none.
 doi:10.1038/437045a

GEOPHYSICS

A moving fluid pulse in a fault zone

In the Gulf of Mexico, fault zones are linked with a complex and dynamic system of plumbing in the Earth's subsurface. Here we use time-lapse seismic-reflection imaging to reveal a pulse of fluid ascending rapidly inside one of these fault zones. Such intermittent fault 'burping' is likely to be an important factor in the migration of subsurface hydrocarbons.

Faults have a dual function in that they can act as both an impediment to and, at times, a preferential pathway for fluid flow. Both types of behaviour are invoked in the petroleum industry to explain how hydrocarbons move from the location at which they are generated (for example, by flowing along faults¹) into fault-bounded reservoirs where they become trapped (for example, by a lack of flow across faults²). Several lines of evidence from the South Eugene Island Block 330 field, offshore Louisiana, USA, indicate that faults cutting through sequences of Pliocene and Quaternary sands and shales have hosted significant vertical fluid flow over the past 250,000 years, continuing to the present day^{3–5}.

We present an additional set of data obtained from seismic reflection imaging that

indicates fast fluid movement (more than 100 metres per year) along growth faults. Previously, we have demonstrated that reflections from the fault planes appearing in seismic data from South Eugene Island Block 330 contain information about the distribution of fluid pressures across faults⁶.

The South Eugene Island field is an ideal location for this study. Multiple vintages of seismic-reflection surveys can be interpreted in the context of abundant fluid pressure, geochemical and other data. Normal faults transect the field at dips of about 50° southwest and separate upthrown sediments saturated by highly pressurized fluids from relatively less pressurized downthrown sediments (see supplementary information). The fault zones are typically at the same pressure as the upthrown sediments⁴. However, exceptionally pressurized fluid was encountered in one penetration of a growth fault, the B-fault, in the A10ST well^{3,4}. It was proposed that the isolated pocket of anomalously high fluid pressure in the A10ST well could represent a spatially limited pulse of anomalously pressured fluid⁴.

To test the idea of a moving fluid pulse, or fault burp, we isolated the fault-plane reflections from the B-fault

in images derived from seismic surveys taken in 1985 and 1992 and looked for indications of movement. In Fig. 1, we show reflectivity as a function of position on the fault plane for both sets of data. Patches of high reflectivity, or 'bright spots', are known to be associated with the presence of fluids⁷. The most striking pattern in the fault reflectivity maps is the northeast movement of the areas of highest reflectivity

between 1985 and 1992. This movement, in the updip direction, is to be expected for a fluid pulse ascending the B-fault. Also note the correlation between the area of highest reflectivity in 1992 and the location of the intersection with the A10ST well (Fig. 1b), where highly pressurized fluid in the fault was observed in 1993 (ref. 4).

From the reflectivity maps at the B-fault, we estimate the movement of the fluid pulse to be of the order of 1 km between 1985 and 1992. The observed movement of 1 km is significant compared with typical errors encountered in the processing of seismic data (see supplementary information). Movement of 1 km between 1985 and 1992 corresponds to an average pulse speed of about 140 m yr⁻¹. Such geologically fast fluid flow up a vertically permeable fault agrees with the dynamic-capacity model of fault-bounded reservoirs⁸ and is consistent with a nonlinear fluid-flow model involving pressure-dependent permeability⁹. These fault burps are key to the understanding of fluid-migration mechanisms and fault-zone rheology in the Earth's crust.

Matthew M. Haney*, **Roel Snieder***, **Jon Sheiman**‡, **Steven Losh**†§

*Center for Wave Phenomena and Department of Geophysics, Colorado School of Mines, Golden, Colorado 80401, USA

e-mail: mmhaney@sandia.gov

‡Shell International Exploration & Production, Houston, Texas 77025, USA

§Department of Earth and Atmospheric Sciences, Cornell University, Ithaca, New York 14853, USA

†Present addresses: Sandia National Laboratories, Geophysics Department, PO Box 5800 MS 0750, Albuquerque, New Mexico 87185-0750, USA (M.H.); Department of Chemistry and Geology, Minnesota State University, TN242, Mankato, Minnesota 56001, USA (S.L.).

- Hooper, E. C. D. *J. Petrol. Geol.* **14**, 161–180 (1991).
- Holland, D. S., Leedy, J. B. & Lammlein, D. R. in *Structural Traps III, Tectonic Fold and Fault Traps: AAPG Treatise of Petroleum Geology Atlas of Oil and Gas Fields* (eds Beaumont, E. A. & Foster, N. H.) 103–143 (AAPG, Tulsa, 1990).
- Anderson, R. N., He, W., Hobart, M. A., Wilkinson, C. R. & Roice, H. R. *The Leading Edge* **10**, 12–17 (1991).
- Losh, S., Eglinton, L., Schoell, M. & Wood, J. *AAPG Bull.* **83**, 244–276 (1999).
- Whelan, J. K., Eglinton, L., Kennicutt, M. C. II & Qian, Y. *Geochim. Cosmochim. Acta* **65**, 3529–3555 (2001).
- Haney, M. et al. in *Proc. EAGE Spec. Sess. Fault and Top Seals* (eds Engelder, T., Konstanty, J. & Grauls, D.) 0–7 (EAGE, Houston, 2004).
- Dobrin, M. B. *Introduction to Geophysical Prospecting* (McGraw-Hill, New York, 1976).
- Finkbeiner, T., Zoback, M. D., Stump, B. & Flemings, P. *AAPG Bull.* **85**, 1007–1031 (2001).
- Rice, J. R. in *Fault Mechanics and Transport Properties of Rocks* (eds Evans, B. & Wong, T.-F.) 475–503 (Academic, San Diego, 1992).

Supplementary information accompanies this communication on Nature's web site.

Competing financial interests: declared none.
 doi:10.1038/437046a

BRIEF COMMUNICATIONS ARISING online
 ♦ www.nature.com/bca see Nature contents.

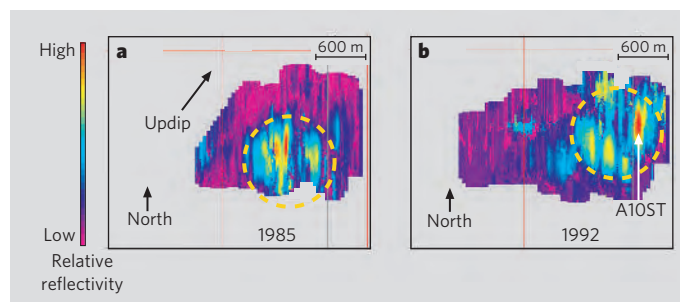


Figure 1 | A fault caught in the act of burping. **a**, Map of the B-fault showing reflectivity from the fault plane in 1985. The area of highest reflectivity is circled in gold. **b**, Map of the B-fault reflectivity, as shown in **a**, but from 1992. The data extend over a slightly larger area than in **a**; however, the spatial perspective is identical. The area of highest reflectivity, circled in gold, is shifted roughly 1 km north-east in the updip direction relative to its location in 1985, as would be expected for a fluid pulse ascending the B-fault; this movement is depicted by the arrow in **a**. Also shown is the location of the A10ST well intersection, where exceptionally high fluid pressures were encountered while drilling into the B-fault zone in 1993.

PLANT GENETICS

Hothead healer and extragenomic information

Arising from: S. J. Lolle, J. L. Victor, J. M. Young & R. E. Pruitt *Nature* **434**, 505–509 (2005)

Lolle *et al.* suggest that non-mendelian inheritance in *Arabidopsis thaliana* might be attributable to an ancestral RNA-sequence cache¹, whereby the RNA genome of previous generations causes a high rate of reversion of the plant's mutant *hothead* (*hth*) and *erecta* (*er*) genes. Here I describe a 'distributed genome' model that also explains their results, in which mutant *hth* DNA is restored by homologous sequences present in the genome itself. This model has implications for the generation of diversity without mating.

DNA-homology searches of the *Arabidopsis* genome based on the 21 nucleotides surrounding *hth-4*, *hth-8*, *hth-10* and *er* reveal the presence of short, perfectly homologous DNA stretches (known as 'reverting sequences') that include nucleotides needed to correct these mutations (Fig. 1). There are also many examples of short homology in genes tested for polymorphism (including *GL1*, *UFO* and *GAPC*).

These sequences might be transcribed into short RNA molecules directed against other chromosomal loci by the cellular machinery, perhaps with the involvement of *DRD1* (ref. 2), producing short RNA–DNA hybrids with potential mismatches that can be corrected by mismatch repair³. Consistent with this model, the *hth-4* allele — with 6 reverting sequences of 13–15 nucleotides each — has a lower reversion frequency than *hth-10*, which has 24 reverting sequences of 13–18 nucleotides. As a result of such differences in number, as well as differences in the production of short RNA, some sequences might be changed more than others.

These short sequences should also produce forward mutations, so it is important to measure forward-mutation frequency for several loci. Also, of the several short sequences available for reversion, some might express more short RNA in the male gamete, explaining the preferential transmission of reverted alleles through pollen. The messenger RNA of the corresponding gene may competitively hybridize with a small length of RNA and prevent its interaction with DNA but, owing to the shortness of the base-paired sequence, the mRNA would not be totally inactivated and so would not produce a mutant phenotype. As both non-sense and missense alleles can cause a reduction in transcription compared with the wild-type allele⁴, there would be more opportunity for reversion to wild type than for forward mutation, which would account for the shielding of the genome against forward mutations.

Conversion to neutral alleles could also occur by this mechanism, but again some sequence stretches might be more effective than others. Lolle *et al.* did not find any neutral

HTH	ATTCGGCCGTCTCACACCGC
<i>hth-4</i>	ATTCGGCCGTCTCACACCGC
RS	TCGGCCGTCTCACAA
HTH	CGAGTCTCCAAGAACCAACCC
<i>hth-8</i>	CGAGTCTCCAAGAACCAACCC
RS	GTCTCCAAGAACCAA
HTH	CAGACTGTTGAATTACAAAG
<i>hth-10</i>	CAGACTGTTGAATTACAAAG
RS	AGACTGTTGAATTACAA
ER	TATGCTTCTTAAGC
<i>er</i>	TATGCTTCTTAAGC
RS	TATGCTTCTTAAGC

Figure 1 | DNA nucleotide sequences of the *hth-4*, *hth-8*, *hth-10* and *er* mutants in the region of the mutation, compared with wild type. Sequences of the mutants are shown in blue, with the mutated nucleotide in lower-case; the corresponding wild-type sequences are in black and the nucleotide at the site of mutation is highlighted in red. Homologous sequences that might cause the mutations to revert (RS sequences), obtained by BLAST-searching the *Landsberg erecta* database from www.arabidopsis.org, are shown in green, with the wild-type nucleotide in red. Reversion frequency is lower for the *hth-4* allele (with 6 RS sequences of 13–15 nucleotides), than for *hth-8* (with 20 RS of 13–15 nucleotides) and *hth-10* (with 24 RS and 13–18 nucleotides).

mutation in nine reverted *HTH* genes¹. Even if the activity of all sequence stretches were comparable, the active reversion at any site should be independent of events at other sites: therefore, the frequency of a neutral mutation among revertants is expected to be around 1%, lower than the level of detection in Lolle *et al.*¹.

The proposed sequence-mediated reversion

frequency could be boosted by another mechanism not described by Lolle *et al.*¹. The *hth* mutant cuticles have increased cellular permeability compared with the wild type⁵. I suggest that the *hth* embryo sac is also more porous than the *HTH* embryo sac, causing DNA in the *HTH/hth* heterozygote to enter the *hth* embryo sac from the two degraded *HTH* spores and become 'archived', as in the P22 phage^{6,7}. Although the increased cellular permeability would allow the *hth* gametophyte to obtain *HTH* molecules in the F₁ generation, in subsequent generations these DNAs need not replicate; the endogenous reverting sequence might provide a basal level of reversion.

Lolle *et al.* propose that metabolic stress in *hth* increases its reversion frequency¹, which might increase information transfer between selected short sequences to alter DNA and create genetic diversity.

Abdel Chaudhury

CSIRO Plant Industry, GPO Box 1600, ACT 2601, Australia

e-mail: abdul.chaudhury@csiro.au

1. Lolle, S. J., Victor, J. L., Young, J. M. & Pruitt, R. E. *Nature* **434**, 505–509 (2005).
2. Matzke, M. A. & Birchler, J. A. *Nature* **455**, 24–34 (2005).
3. Watson, J. D., Losick, R., Bell, S. P., Gann, A. & Levine, M. *Molecular Biology of the Gene* (Cummings, San Francisco, 2003).
4. Mustajoki, S., Kauppinen, R., Mustajoki, P., Suomala, A. & Peltonen, L. *Genome Res.* **7**, 1054–1060 (1997).
5. Lolle, S. J., Hsu, W. & Pruitt, R. E. *Genetics* **149**, 607–619 (1998).
6. Downs, D. M. & Roth, J. R. *Genetics* **117**, 367–380 (1987).
7. Ray, A. *Nature* doi:10.1038/nature04063 (2005).

doi:10.1038/nature04062

Reply: Lolle *et al.* reply to this communication (doi:10.1038/nature04064).

PLANT GENETICS

RNA cache or genome trash?

Arising from: S. J. Lolle, J. L. Victor, J. M. Young & R. E. Pruitt *Nature* **434**, 505–509 (2005)

According to classical mendelian genetics, individuals homozygous for an allele always breed true. Lolle *et al.*¹ report a pattern of non-mendelian inheritance in the *hothead* (*hth*) mutant of *Arabidopsis thaliana*, in which a plant homozygous at a particular locus upon self-crossing produces progeny that are 10% heterozygous; they claim that this is the result of the emerging allele having been reintroduced into the chromosome from a cache of RNA inherited from a previous generation. Here I suggest that these results are equally compatible with a gene conversion that occurred through the use as a template of

DNA fragments that were inherited from a previous generation and propagated in archival form in the meristem cells that generate the plant germ lines. This alternative model is compatible with several important observations by Lolle *et al.*¹.

Such archival forms of DNA have been described previously². The template DNA could have originated in the fragmented genomes of three of the four haploid female meiotic products in the germline of a heterozygous plant. Within the ovule of the original heterozygous plant, the surviving haploid female germline cell containing the mutant *hth*

allele acquires from the degenerating sister nuclei a collection of chromatin fragments that are presumably heterochromatinized and silenced; they might also be covalently modified and therefore hard to detect by Southern blotting or by amplification in the polymerase chain reaction.

I propose that these supernumerary chromatin fragments propagate within the meristem cells of succeeding generations, and so are present in a very few cells of the plant; nevertheless, they are often present in the germ lines that are themselves derived from the meristem. Within a second-generation descendant that is homozygous for a chromosomal allele, such supernumerary chromatin fragments might harbour another allele in a cryptic state. These supernumerary fragments might pair with normal chromosomes in the male germ line preferentially, as indicated by Lolle *et al.*¹, where they direct gene conversion of the chromosomal alleles. The gene-converted allele, therefore, would reappear as a mendelian factor in the third or a subsequent generation.

This proposal dispenses with the hypothetical RNA cache¹, for which little evidence exists. It is consistent with the incidence of supernumerary chromosome fragments in

plants, particularly with their known effects on pollen function³, and explains the lack of somatic convertant sector in any generation¹; it also invokes standard molecular processes of DNA-directed gene conversion over relatively long regions of homology. The differences in the frequency of conversion of the three tested alleles¹ may reflect co-conversion polarity⁴.

My proposal can also explain the curiously high frequency of allele reappearance in dissected embryos, which is roughly twice that found in mature plants (it increased from 10% to 20%). Assume that gene conversion produces a heteroduplex DNA that escapes mismatch repair. If conversion occurs in the haploid generative nucleus of the pollen, the two sperm cells will be non-identical at the converted allele. After double fertilization, the embryo and the endosperm will contain two different alleles. An embryo dissected from such a seed will inevitably be associated with endosperm cells, and will yield a signal in the polymerase chain reaction that is indistinguishable from a heterozygous embryo, although the embryo itself is homozygous for the allele. As this will occur as often as a conversion, the frequency of the total converted allele in dissected embryos and endosperms together should be

about 20%, an estimate close to the experimentally observed rate.

The model proposed here, but not one in which gene conversion is directed by RNA⁵ or by very short, dispersed, repeated DNA sequences⁶, is easily reconciled with the notion of co-conversion polarity⁴. It should be possible to test whether co-conversion polarity is a factor in the phenomenon revealed by Lolle *et al.*¹ by producing double- and triple- mutant alleles through exploitation of this effect. Further investigations into the possibility and nature of DNA archiving in plants and of plant germ lines should prove interesting.

Animesh Ray

Keck Graduate Institute, Claremont, California 91711, USA

e-mail: aray@kgi.edu

1. Lolle, S. J., Victor, J. L., Young, J. M. & Pruitt, R. E. *Nature* **434**, 505–509 (2005).
2. Downs, D. M. & Roth, J. R. *Genetics* **117**, 367–380 (1987).
3. Stebbins, G. L. in *Chromosomal Evolution in Higher Plants* 67–71 (Addison-Wesley, Reading, Massachusetts, 1971).
4. Nicolas, A. & Petes, T. D. *Experientia* **50**, 242–262 (1994).
5. Baltimore, D. *Cell* **40**, 481–482 (1985).
6. Chaudhury, A. *Nature* doi:10.1038/nature04062 (2005).

doi:10.1038/nature04063

Reply: Lolle *et al.* reply to this communication (doi:10.1038/nature04064).

PLANT GENETICS

Lolle et al. reply

Reply to: A. Chaudhury (doi:10.1038/nature04062) and A. Ray (doi:10.1038/nature04063)

Chaudhury¹ and Ray² propose alternative models to account for our observed pattern of non-mendelian inheritance in the *hothead* (*hth*) mutant of *Arabidopsis*³.

Chaudhury suggests that the information required to restore correct genetic sequences in *hth* mutant plants could be stored in short stretches of nucleotide sequence within the genome¹. Although the sequences required for restoration are indeed present in the genome, the length of similarity seen in the 'reverting sequences' identified by Chaudhury is barely greater than would be expected from random chance. An appropriate control for his *in silico* experiment would be to establish how many similar sequences (13–18 nucleotides in length, with a single nucleotide mismatch relative to the sequence in the parent plant) are present in the genome that could likewise introduce silent nucleotide substitutions into the *hth* gene under the same conditions.

Should there be a significant number of these sequences (and given that no such silent mutations occur in the corrected alleles³), then an explanation is needed for why some of them are used for correction whereas the majority are not. Even if silent mutation events occur independently of the reversion of the *hth* mutation, we should still detect them in our small sample of sequenced reverted alleles

owing to the relatively large number of possible silent mutations.

Chaudhury also suggests that the increased permeability of mutant *hth* female gametophytes could allow DNA or RNA from the degenerating non-functional megaspores to enter the functional *hth* megaspore and then be archived and carried forward to allow gene conversion in the next generation. This may be a possibility, but our previous work examined only changes in the permeability of the extracellular cuticle covering the outside of the epidermal cell layer⁴. We have no data concerning increased cellular permeability in *hth* mutants. The fact that we see no obvious drop in the rate of reversion over several generations³ is inconsistent with Chaudhury's suggestion that this could be a second mechanism to bolster the rate of reversion for only a single generation.

Ray² claims that our results could be explained by the stable inheritance of supernumerary chromosomal fragments that are archived in a way that makes them inaccessible to DNA hybridization and the polymerase chain reaction. These fragments might also be restricted to meristematic cells and therefore be present in such low concentrations that they are undetectable in a conventional experiment. This is an interesting possibility that is consistent with our observations, but it postu-

lates a novel system of segregation to restrict the chromosome fragments to what would constitute a hitherto undetected germ line in plants. Considering also Ray's explanation for the doubled rate of conversion in embryos, we note that it would be necessary for all conversion events to take place in the generative cell and to fail to be corrected by mismatch repair.

In summary, we agree with Ray that there is little direct evidence to support any given molecular identity for the cryptic templates that allow genetic restoration in *hth* mutant plants. We proposed that the templates might be a replicating form of RNA, but the data are also consistent with a form of DNA that is segregated into a limited number of cells in the plant or that is not readily detectable by conventional molecular techniques. This sequence archive (whether DNA or RNA) would therefore require the same basic properties as those we proposed³: it would need to be replicated, transmitted with high fidelity over several generations, and retain the ability to restore nuclear DNA sequences.

Susan J. Lolle, Jennifer L. Victor,

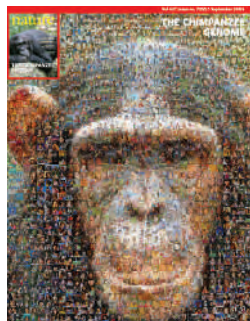
Jessica M. Young, Robert E. Pruitt

Department of Botany and Plant Pathology, Purdue University, West Lafayette, Indiana 47907-2054, USA

e-mail: pruittr@purdue.edu

1. Chaudhury, A. *Nature* doi:10.1038/nature04062 (2005).
2. Ray, A. *Nature* doi:10.1038/nature04063 (2005).
3. Lolle, S. J., Victor, J. L., Young, J. M. & Pruitt, R. E. *Nature* **434**, 505–509 (2005).
4. Lolle, S. J., Hsu, W. & Pruitt, R. E. *Genetics* **149**, 607–619 (1998).

doi:10.1038/nature04064



THE CHIMPANZEE GENOME

The mosaic on the cover to this special section is by Leslie Gaffney and Eric Lander, created by Runaway Technology, Inc. with PhotoMosaic by Robert Silvers. (Courtesy of the Broad Institute of Harvard and MIT.) The issue cover shows an adult female chimpanzee, Jolie of the Ngogo community in Kibale National Park, Uganda, less than a month before she gave birth to her first infant. The photo is part of photographer Kevin Langergraber's ongoing study of the effect of genetic relatedness on patterns of affiliation and cooperation in wild chimpanzees.

Until now, genome sequence information has shown us how many seemingly very different organisms are amazingly like humans. At a conservative estimate we share about 88% of our genes with rodents and 60% with chickens. Applying a more liberal definition of similarity, up to 80% of the sea-squirt's genes are found in humans in some form. So it's no surprise that we are still asking, "What makes us human?" To apply genomics to this quest, we need to shift the focus to look at our closest living relative, the chimpanzee. Given that we share more than 98% of our DNA and almost all of our genes, chimps are the best starting point to study not the similarities, but the minute differences that set us apart.

We are therefore extremely pleased to present this special section to commemorate the genome of the common chimpanzee, *Pan troglodytes*. In doing so, we hope to provide a resource for more than just genomics. We introduce the section with a timeline that charts the history of the chimp. This is followed by four Progress pieces that review recent work on chimp culture and behaviour, psychology and neural processing of number systems, as well as a closer look at brain anatomy and neurogenetics at the single-gene level.

On page 69, the Chimpanzee Sequencing and Analysis Consortium reports analysis of the long-awaited draft genome sequence. This is supported on page 101 by Hughes *et al.*, with the sequence of part of the chimpanzee Y chromosome. Comparing the genetic code of humans and chimps will allow us to comb through each gene or regulatory region to find single changes that might have made a difference in evolution, and the authors list some new candidates for further study. Two more research papers by Cheng *et al.* (page 88) and Linardopoulou *et al.* (page 94) detail changes in highly variable regions in the human and chimp genomes; additions or deletions of larger chunks of DNA may be as important as single nucleotide changes in shaping our genomes.

Finally, we need physical evidence to tell us how chimps and humans may have lived millions of years ago. Surprisingly, to date there has been no fossil record of the chimp; on page 105, McBrearty and Jablonski report the first unequivocal fossil evidence of the genus *Pan*.

Chris Gunter, Senior Editor
Ritu Dhand, Chief Biology Editor
Commissioning Editors: Tanguy Chouard, Henry Gee, Jane Rees & John Spiro

48 TIMELINE A brief history of chimps

50 NEWS AND VIEWS The chimpanzee and us
 W-H Li & M A Saunders

PROGRESS

52 The second inheritance system of chimpanzees and humans
 A Whiten

56 A century of getting to know the chimpanzee
 F B M de Waal

60 Our chimpanzee mind
 M Hauser

64 Molecular insights into human brain evolution
 R S Hill & C A Walsh

ARTICLES

69 Initial sequence of the chimpanzee genome and comparison with the human genome
 The Chimpanzee Sequencing and Analysis Consortium

88 A genome-wide comparison of recent chimpanzee and human segmental duplications
 Z Cheng, M Ventura, X She, P Khaitovich, T Graves, K Osoegawa, D Church, P DeJong, R K Wilson, S Pääbo, M Rocchi & E E Eichler

94 Human subtelomeres are hot spots of interchromosomal recombination and segmental duplication
 E V Linardopoulou, E M Williams, Y Fan, C Friedman, J M Young & B J Trask

LETTERS

101 Conservation of Y-linked genes during human evolution revealed by comparative sequencing in chimpanzee
 J F Hughes, H Skaletsky, T Pyntikova, P J Minx, T Graves, S Rozen, R K Wilson & D C Page

105 First fossil chimpanzee
 S McBrearty & N G Jablonski

A brief history of chimps

As our closest living relative, the chimpanzee holds a unique fascination for researchers from anthropologists to geneticists. Here, we chart the history of mankind's relationship with the chimp, culminating in this week's landmark genomic achievement.

5 MILLION TO 7 MILLION YEARS AGO Origin of species

The last common ancestor of humans and chimpanzees is believed to have walked on four legs. The oldest fossils that resemble bipedal humans are 6 million–7 million years old, although DNA sequence analyses suggest the two lineages separated about 5 million years ago.

1 MILLION TO 2 MILLION YEARS AGO



Double the fun

The chimpanzee (*Pan troglodytes*) diverges from the lineage leading to the bonobo (*Pan paniscus*), or pygmy chimpanzee.

1641



Is this a chimp I see before me?

Nicolaas Tulp, a Dutch anatomist, is the first formally to describe an ape, although it is not clear from historical records whether his subject was a chimpanzee, a bonobo or an orang-utan.

1699

More than skin deep

Edward Tyson, an English physician, publishes his account of the first confirmed dissection of a chimpanzee. He echoes remarks made by Tulp, pointing out that chimp anatomy is remarkably similar to our own.



1739

A true likeness

French artist Louis Gérard Scotin draws a chimp — possibly the first one imported live into Europe.

1775

Going underground?

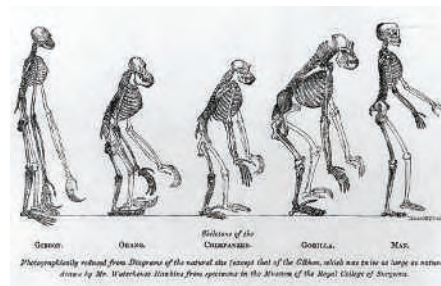
The species name *troglodytes* is coined by Johann Friedrich Blumenbach, a German anthropologist. It is derived from the name for an African race of cave dwellers — perhaps mythical — from the middle ages. Why he chose this name remains unclear.

1816

From god to genus

German naturalist Lorenz Oken is the first to use *Pan* as the name for the genus to which chimpanzees belong — named after the hairy, Greek rural god.

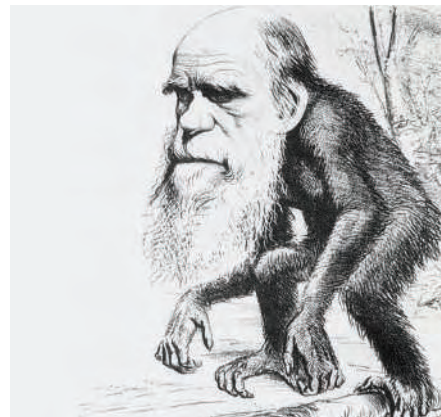
1863



Order, order!

Humans should be placed in the same order as apes, argues Thomas Henry Huxley.

1871



The Descent of Man

In his follow-up to *On the Origin of Species*, Charles Darwin suggests that humans are most closely related to African great apes.

1917

No problem

German psychologist Wolfgang Köhler describes how captive chimpanzees solve problems, using boxes and sticks to retrieve bananas hanging from ropes in their enclosures.



1961

Space shot

The four-year-old chimpanzee Ham is the first chimpanzee in space, experiencing more than six minutes of weightlessness and moving levers in response to flashing lights.

1964

JANE GOODALL INST.

**Chimp technology**

Jane Goodall, in a study published in *Nature*, shows that wild chimpanzees make and use tools. She watches them fashion twigs into implements to catch termites.

1967

Close cousins

Vincent Sarich and Allan Wilson argue in *Science* that African great apes are our closest relatives, based on studies of blood proteins. They find that antibodies against human proteins cross-react best with chimpanzee and gorilla proteins.

1970

S. SAVAGE-RUMBAUGH

**Hello, handsome**

Chimpanzees are shown to recognize themselves in mirrors, which Gordon Gallup interprets as evidence of a level of self-awareness. Subsequent research has demonstrated 'mirror self-recognition' in bonobos, gorillas and orang-utans, but not in other primates.

1978

Pass the banana

Sue Savage-Rumbaugh and her colleagues publish a study in *Science* on abstract communication skills: two chimpanzees learn symbols and use them to request food from one another.

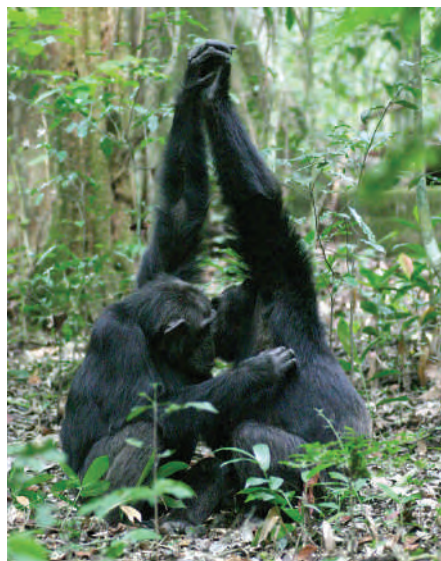
F. LANTING/FLPA



1978

When in Mahale...

William McGrew and Caroline Tutin are the first to describe a 'social custom' among the chimpanzees of Mahale, Tanzania. They perform a 'grooming hand-clasp' — behaviour absent at Gombe, only 100 kilometres away.



K. LANGERABER/UNIV. MICHIGAN

1984

Pairing up

Charles Sibley and Jon Ahlquist study how easily DNA strands of humans and chimpanzees form 'complementary' pairs. The experiments suggest that the two species' genomes are 98.4% similar.

1997

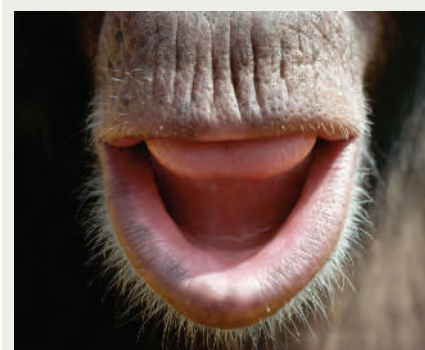
Let's get started

Edwin McConkey and Morris Goodman make the first published call to compare the genomes of chimpanzees and humans, in what they call the Human Genome Evolution Project.

1998

Vive le difference!

In the first detailed study on genetic differences — as opposed to similarities — between humans and chimpanzees, Ajit Varki's team shows that part of the gene for CMP-sialic acid hydroxylase is missing in humans, but not in chimpanzees. The enzyme converts one form of a cell-surface molecule into another.



Y. ARTHUS-BERTRAND/CORBIS

AUGUST 2002

Speaking in tongues

Svante Pääbo's group reports in *Nature* that *FOXP2*, a gene important in speech articulation and other linguistic skills, is different in chimpanzees and humans.

FEBRUARY 2002

Making the case

White papers are submitted to the US National Human Genome Research Institute (NHGRI) arguing for the chimpanzee genome to be sequenced.

MAY 2002

The starting gun

Sequencing the chimpanzee genome is declared a priority by the NHGRI.

OCTOBER 2002

Not so similar

Roy Britten is the first to include deletions and insertions when comparing the genomes of humans and chimpanzees. He argues that the two species share 95% of their DNA, less than was thought, although the figure remains controversial.

DECEMBER 2003

First draft

The NHGRI announces an initial aligned draft sequence of the chimpanzee genome.

MAY 2004

One down...

A Japanese consortium publishes in *Nature* the sequence of chimpanzee chromosome 22, the counterpart to human chromosome 21.

SEPTEMBER 2005

The chimp genome

A draft sequence of the chimpanzee genome is published in this issue of *Nature*.



NEWS & VIEWS



The chimpanzee and us

Wen-Hsiung Li and Matthew A. Saunders

Publication of the draft DNA sequence of the chimpanzee genome is an especially notable event: the data provide a treasury of information for understanding human biology and evolution.

What genetic changes make us so different from the chimpanzee, our closest relative? Scientists have been trying to answer this challenging question for decades, and publication of the draft of the chimpanzee genome (page 69 of this issue)¹ is a significant step forward. The species studied is the common chimpanzee, *Pan troglodytes*; its only 'sister' species is the pygmy chimpanzee or bonobo, *Pan paniscus* (Fig. 1).

The draft tells us that the DNA sequence of our genome and that of the chimpanzee differ by only a few per cent. This still amounts to tens of millions of differences because each genome contains some 3 billion nucleotides. One way to determine what the important differences are is to identify evolutionary changes that are specific to us, *Homo sapiens*. Another is to look for signatures of positive natural selection in the sequences of the two genomes. Both of these approaches, and other comparative analyses, are described in the draft-genome paper¹ and the companion papers (pages 88–104)^{2–4}.

The assembly of a complete genome requires multiple rounds of sequencing. The chimpanzee genome draft represents a sequencing coverage of about 3.5 times, lower than that in the initial publication of other genomes, such as those of human, mouse and rat. Nonetheless, the draft is extremely useful for showing general differences between the chimpanzee and human genomes. The new data show that they differ by only 1.23% in terms of nucleotide substitutions. This is identical to a previous estimate from a mere 53 regions, each of about 500 base pairs, randomly chosen from the genome⁵.

The sequence divergence varies among genomic regions, presumably because of regional variations in mutation rate, selective constraints and the rate of sequence exchange (recombination) between chromosome pairs during cell division. The highest divergence is found for the Y chromosome and the lowest for the X chromosome. This is expected, because the Y chromosome is present only in males, which have a higher germ-line mutation rate than females, whereas the X chromosome is carried in both females and males.

Natural selection is commonly thought to operate mainly at the protein level. For this

reason, nucleotide changes in protein-coding regions are usually classified into two groups: 'synonymous changes' (which do not cause any change in amino acids) and 'non-synonymous changes' (which do cause amino-acid changes). If a coding region is subject to strong selective constraints, then the non-synonymous substitution rate (K_A) will be considerably lower than the synonymous substitution rate (K_S); that is, the K_A/K_S ratio will be less than 1. On the other hand, if a gene is subject to very weak selective constraints or continued positive selection, K_A/K_S may be close to 1 or even higher.

Comparison¹ of 13,454 human–chimpanzee gene pairs gives an average K_A/K_S of 0.23, much lower than previously estimated from more limited data sets of human–chimpanzee (0.63)⁶ and human–baboon (0.34)⁷ comparisons. This ratio is twice that estimated from the mouse–rat comparison (0.13): this is probably due to less effective purifying selection, a process that eliminates deleterious mutations, in species with relatively small population sizes such as primates. Importantly, the new estimate is similar to the K_A/K_S from data on variation among humans (~0.20–0.23), suggesting that the proportion of advantageous mutations along the human lineage is lower than previously estimated^{7,8}. A total of 585 genes (more than that expected at random) do, however, display a higher K_A than the substitution rate in non-coding sequence

(K_I). The highest K_A/K_I examples include the genes that encode glycoporphin C, granulysin, protamine and semenogelin, proteins that are involved in immunity or reproduction.

Duplications, insertions and deletions

Although single-nucleotide substitutions are commonly considered when quantifying sequence divergence, insertions/deletions (indels) and recent duplications of DNA segments account for a markedly larger proportion of the difference between the human and chimpanzee genomes (3% and 2.7%, respectively). More than a third of the indels are due to repeated sequences, and about a quarter to transposable elements. These are DNA sequences that can move to different genomic regions, two of the major classes being Alu elements (short transposable sequences about 300 base pairs long) and L1 elements (long transposable sequences).

There are approximately 7,000 Alu elements in the human genome but only about 2,300 in the chimpanzee genome, indicating that these elements have been less active in the chimpanzee. L1 elements, however, have been equally active in the two genomes — against the previous estimate of two- to three-fold higher activity in the chimpanzee⁹. The functional importance, if any, of these differences remains unknown. Recent segmental duplications (of longer than 20 megabases and

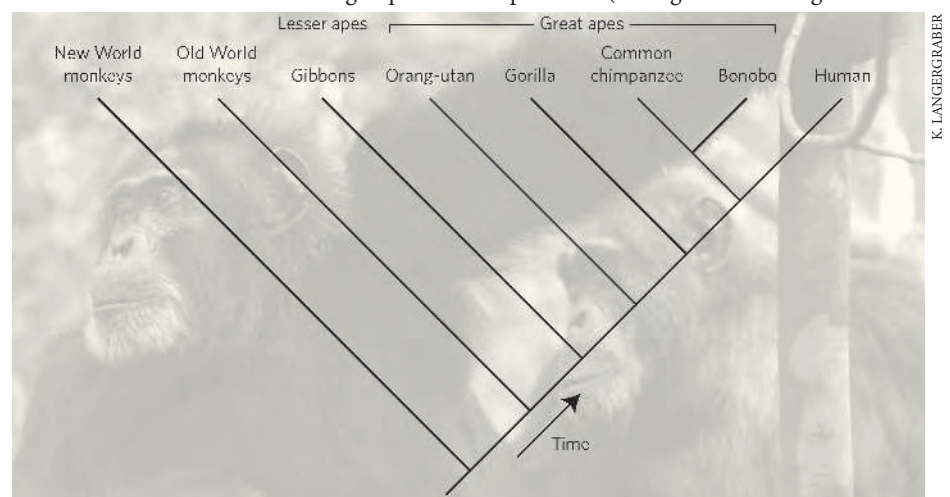


Figure 1 | Evolutionary relationships among the higher primates. Divergence of the chimpanzee and human lineages occurred about 6 million years ago; the times of lineage divergence are not to scale.

greater than 94% sequence identity) are common in both genomes². But although about 33% of human duplicated segments are human-specific, only about 17% of chimpanzee duplicated segments are chimpanzee-specific. Interestingly, about half of the genes in the human-specific duplicated regions exhibit significant differences in gene expression relative to the chimpanzee, and are most often upregulated.

Human genetic variation

The chimpanzee genome places the wealth of data on existing genetic variation in humans into evolutionary context. It now becomes possible to determine the ancestral states of that variation, and, with the aid of gene-frequency data in human populations, we may uncover 'footprints' of positive selection that occurred recently (less than 250,000 years ago, say) in humans. Under selective neutrality, new variants should rarely be found at high frequency, and between-species divergence should be correlated with the level of within-species genetic variation. The current analyses identify only six genomic regions that display significantly less variation than expected from the divergence between the *Homo* and *Pan* lineages, which split about 6 million years ago; each of these regions suggests the recent action of positive selection in humans. The power of such a method will increase substantially with the completion of genome drafts of a more distantly related primate such as an Old World monkey or the orang-utan, both of which are in progress¹⁰ (Fig. 1; see also page 17).

What makes us human?

The question of what genetic changes make us human is far more complex. Although the two genomes are very similar, there are about 35 million nucleotide differences, 5 million indels and many chromosomal rearrangements to take into account. Most of these changes will have no significant biological effect, so identification of the genomic differences underlying such characteristics of 'humanness' as large cranial capacity, bipedalism and advanced brain development remains a daunting task. Given the short time since the human–chimpanzee split, it is likely that a few mutations of large effect are responsible for part of the current physical — phenotypic — differences that separate humans from chimpanzees and other great apes.

There are three prevailing hypotheses to account for the evolution of 'humanness traits': protein evolution, the 'less-is-more' hypothesis¹¹, and changes in the regions of the genome that regulate gene activity¹² (Fig. 2). Preliminary analyses of the human and chimpanzee genomes provide some clues about the relative contributions of these effects.

First, consider protein evolution. Are amino-acid changes that have contributed to 'humanness' to be found in rapidly evolving proteins? Most of those genes that do show a

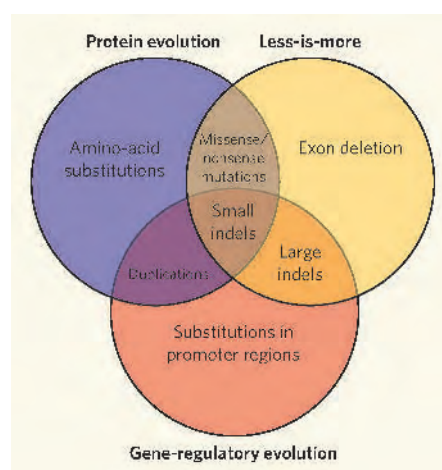


Figure 2 | Hypotheses to explain the genetic underpinnings of human-specific traits. Each of the three hypotheses — protein evolution, 'less-is-more', and gene-regulatory evolution — is depicted by a circle, with note of the mechanisms or processes that could underlie the evolutionary change. A missense mutation causes an amino-acid change; a nonsense mutation causes a sense codon to change into a stop codon, resulting in premature termination of DNA transcription. Indels are insertions/deletions of DNA segments; exons are coding sequences; promoter regions regulate gene activity in various ways.

K_A/K_S of more than 1 are not involved in processes related to supposed humanness traits. In fact, genes related to brain function and neuronal activity show lower-than-average K_A/K_S values. The genes that display high K_A/K_S are mostly related to host–pathogen interaction, immunity and reproduction. This pattern is also found in rats, mice and other mammals. This suggests that protein evolution may not be a major contributor to the evolution of traits unique to humans. But before dismissing this possibility, we must bear in mind that the K_A/K_S test is biased towards genes that experience repeated amino-acid replacements. Genes involved in immunity and reproduction are particularly affected by these processes. But a gene that experiences a 'selective sweep' as a result of only a few changes — because those changes are strongly advantageous — would not leave a significant signal on K_A/K_S . For example, two amino-acid changes alone in the highly conserved FOXP2 protein, a gene-transcription factor, might have contributed to the human capacity for speech¹³. Finally, the role of indels and gene duplications in human–chimpanzee protein evolution remains largely unexplored.

Second, the 'less-is-more' hypothesis posits that loss-of-function changes relative to the 'prototypical ape' traits are characteristic of certain humanness traits — for example, lack of body hair, preservation of some juvenile traits into adulthood and expansion of the cranium. Such loss-of-function changes could be caused by non-synonymous substitutions, indels, loss of coding regions and deletion of entire genes. The comparisons to the chim-

panzee have unveiled 53 human genes with disruptive indels in the coding regions, and genes in this category may be associated with intriguing phenotypes^{14–16}. Indels could plausibly be major contributors to human–chimpanzee phenotypic differences, especially given that these mutations can also influence the two other proposed mechanisms for the evolution of humanness (Fig. 2).

Third, there is the long-standing hypothesis that the phenotypic differences between humans and chimpanzees primarily arise from changes in gene-regulatory regions. The current analyses¹ do not address this issue in detail, because it is still notoriously difficult to identify such regions. Most of our current knowledge about regulatory regions comes from identifying similarities between distantly related species. The matter could be addressed further in a comparative genomic framework by identifying conserved regulatory regions among relatively closely related species¹⁷, including Old World monkeys, in conjunction with a comparison to the chimpanzee sequence and with microarray expression studies that can provide functional validation. The hypothesis invoking evolution in gene-regulating regions is currently the hardest to test. Yet it may be the most promising, given what we know of human biology relative to that of apes.

The draft of the chimpanzee genome is an exciting addition to the list of sequenced vertebrate genomes. Next to the human genome itself, it is the most useful for understanding human biology and evolution. But the data still leave many questions unanswered about what genetic modifications underlie the major features distinguishing *Homo sapiens* from the great apes. The next stages of this grand project will involve finer-scale investigation of individual regions and genes to reveal the details of the general patterns now uncovered at the genomic level.

Wen-Hsiung Li and Matthew A. Saunders are in the Department of Ecology and Evolution, University of Chicago, Chicago, Illinois 60637, USA. e-mail: whli@uchicago.edu

1. The Chimpanzee Sequencing and Analysis Consortium *Nature* **437**, 69–87 (2005).
2. Cheng, Z. et al. *Nature* **437**, 88–93 (2005).
3. Linardopoulou, E. V. et al. *Nature* **437**, 94–100 (2005).
4. Hughes, J. F. et al. *Nature* **437**, 101–104 (2005).
5. Chen, F. C. & Li, W.-H. *Am. J. Hum. Genet.* **68**, 444–456 (2001).
6. Eyre-Walker, A. & Keightley, P. D. *Nature* **397**, 344–347 (1999).
7. Fay, J. C., Wyckoff, G. J. & Wu, C. I. *Genetics* **158**, 1227–1234 (2001).
8. Clark, A. G. et al. *Science* **302**, 1960–1963 (2003).
9. Mathews, L. M., Chi, S. Y., Greenberg, N., Ovchinnikov, I. & Swergold, G. D. *Am. J. Hum. Genet.* **72**, 739–748 (2003).
10. <http://www.genome.gov/10002154>.
11. Olson, M. V. & Varki, A. *Nature Rev. Genet.* **4**, 20–28 (2003).
12. King, M. C. & Wilson, A. C. *Science* **188**, 107–116 (1975).
13. Enard, W. et al. *Nature* **418**, 869–872 (2002).
14. Stedman, H. H. et al. *Nature* **428**, 415–418 (2004).
15. Hahn, Y. & Lee, B. *Bioinformatics* **21**, 1186–1194 (2005).
16. International Human Genome Sequence Consortium. *Nature* **431**, 931–945 (2004).
17. Boffelli, D. et al. *Science* **299**, 1391–1394 (2003).

PROGRESS

The second inheritance system of chimpanzees and humans

Andrew Whiten¹

Half a century of dedicated field research has brought us from ignorance of our closest relatives to the discovery that chimpanzee communities resemble human cultures in possessing suites of local traditions that uniquely identify them. The collaborative effort required to establish this picture parallels the one set up to sequence the chimpanzee genome, and has revealed a complex social inheritance system that complements the genetic picture we are now developing.

In the first decade of the twenty-first century, we find ourselves at a point that future generations will probably reflect on as unique in the scientific relationship between humans and chimpanzees. In historical—let alone evolutionary—time frames, this window has only just opened. We have progressed from a position of almost complete ignorance about wild chimpanzees just decades ago¹, to having gathered very detailed knowledge through hundreds of field and laboratory studies^{2–9}. Tragically, however, just as new levels of integrative analysis appear on the horizon, including comparative genomics¹⁰ and the parallel analysis of culture outlined below, the scientific window so recently opened has already started to close as the human species inexorably exterminates its closest relatives. The very material of interest—genetic and cultural diversity—is melting away¹¹. This state of affairs adds urgency to our attempts to grasp the big picture of what it means to be a chimpanzee, and thus, by comparison, what it means to be the human ape.

When we focus our comparative lens on culture, the evidence is all around us that a gulf separates humans from all other animals. Nevertheless, recent studies of great apes suggest that they resemble us culturally to an extent unmatched by other species^{8,12–17}. Many species of fish, birds and mammals have been shown to have ‘traditions’, but if we follow those authors who reserve the term ‘culture’ for the more particular manifestations of tradition that

characterize our own species (Box 1), we are increasingly finding marked cultural similarities between ourselves and other apes.

The following sections will examine three aspects of culture that go beyond the existence of tradition *per se*. These are (1) the population-level patterning of traditions, (2) the mechanisms facilitating transmission of traditions, and (3) the specific behavioural content of traditions. Comparison among chimpanzees, humans and other species in each of these dimensions helps to delineate what chimpanzees share with humans and just where the differences begin.

Population-level cultural complexity

As observations accumulated in the 1980s and 1990s, researchers began to compile a growing list of putative traditional variations among wild chimpanzees^{2,18}. More recently, research directors of the longest-running chimpanzee study sites have pooled their hard-earned data to create the first systematic overview of behavioural variation. Identifying no fewer than 39 different traditions^{12,13} (Fig. 1), the collaborative effort of this international project parallels that which made possible the chimpanzee genome project.

A collaborative exercise applying precisely the same systematic method to orangutans subsequently identified 19 clearly defined traditions, with an additional five more tentatively classified¹⁴. In contrast, studies of traditions in other primates, other mammals,

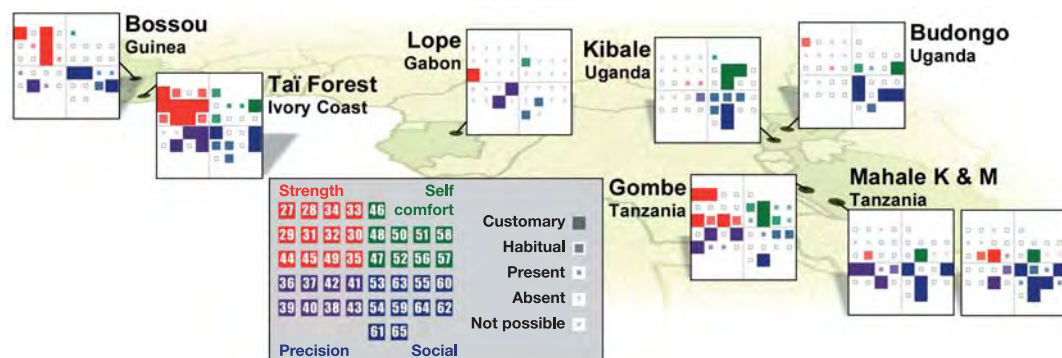


Figure 1 | The cultures of wild chimpanzees. Each chimpanzee community has its own unique array of traditions that together constitute the local ‘culture’. ‘Customary’ acts are those typical in the community, ‘habitual’ ones are less common but consistent with social transmission, and ‘absent’ acts are those missing with no apparent straightforward environmental explanation. Traditions are defined as behaviour patterns that are

customary or habitual in at least one site but absent elsewhere. Transmission is attributed to social learning on the basis of a complex of circumstantial evidence, ranging from intense observation by juveniles to distributions inconsistent with alternative explanations^{12,13,15}. The numbers in cells refer to behaviour patterns in the catalogue of ref. 13, illustrated at <http://culture.st-and.ac.uk/chimp>.

¹Centre for Social Learning and Cognitive Evolution, and Scottish Primate Research Group, School of Psychology, University of St Andrews, St Andrews, Fife KY16 9JP, UK.

birds and fish have most commonly identified no more than a single tradition, and in no such species has the number of behavioural variants reached double figures^{16,17,19,20}. A working hypothesis is thus that in the ancestors of the great ape clade, there occurred a distinctive step towards reliance on a relatively rich cultural repertoire that incorporates both technological and social elements¹⁴.

Figure 1 suggests an additional way in which a distinction between tradition and culture can be made. Each local chimpanzee community has a unique array of specific traditions, representing a 'package' that can be described as its local culture, in the same way that we might distinguish between Scottish and English culture, which are defined by distinctive arrays of traditions. In contrast, studies have yet to identify cultures defined by multiple traditions in species other than primates and cetaceans.

Every year now sees new putative traditions added to the chimpanzee catalogue, and a research consortium, recently formalized as the Collaborative Chimpanzee Cultures Project, Round 2 (CCCP-2), has embarked on a new collation of the accumulated records. Recent investigations address more fine-grained phenomena, such as subtle contrasts in social conventions between neighbouring communities²¹. This is nicely illustrated by recent work on the 'grooming hand-clasp' (Box 2). Data from new study sites have identified additional behavioural variants, such as the recently described 'social scratch': for example, chimpanzees at Ngogo in Uganda will scratch each other's backs using a short jabbing technique, in contrast with the long, raking style observed at Mahale²². Perhaps the most remarkable recent discovery comes from the young

study site at Goulougo, where chimpanzees prepare a two-part tool set for termite fishing²³, unlike the simpler technologies well known at other sites^{12,13} (Box 3).

Ape culture may be particularly complex among non-human animals, yet it clearly falls far short of human culture. An influential contemporary view is that the key difference lies in the human capacity for cumulative culture^{24,25}, whereby the achievements of successive generations have built on previous developments to create complex structures such as languages and technologies. Chimpanzees have accumulated many traditions, but each remains sufficiently simple that there is little scope for it to have developed significant complexity compared to its original form. Hints of cumulation exist¹⁵, such as the refinement of using prop stones to stabilize stone anvils during nut-cracking⁷, but these remain primitive and fleeting by human standards. One possible explanation that has been offered for this human–chimpanzee difference lies in the social learning mechanisms available to each species²⁵, an issue that new genetic approaches based on the complete chimpanzee genome sequence may help to unravel.

The psychological bases of cultural transmission

The circumstantial evidence for traditions among wild chimpanzees is strong, but experiments are required to identify the different social and non-social learning processes involved. Field researchers have been understandably resistant to experimental interventions that might seriously perturb the naturalness of the behavioural ecology they study. Matsuzawa and his colleagues have pioneered subtle field experiments, for example introducing to a community of nut-cracking chimpanzees at Bossou various new nuts that are cracked by chimpanzees at more distant locations^{26,27}. When this introduction coincided with the immigration of a female already expert in dealing with the new (*Coula*) nuts, close observation of this female by other chimpanzees was followed by the gradual spread of *Coula*-cracking to a majority of the community.

Box 1 | Animal traditions, human cultures?

Some authors, particularly in the biological sciences, draw no distinction between the terms 'tradition' and 'culture'¹⁶. Others discriminate between the two in a variety of ways, typically driven by a perceived need to recognize that there is more to culture (human culture in particular) than the existence of tradition alone^{24,25,37}. These terminological differences must be recognized before any sensible comparative analysis can be made of the diverse claims to have identified 'cultural' phenomena across the animal kingdom. Of the two terms, 'tradition' is used with the greatest consistency. Most would probably be content with the definition in a recent survey of the biology of traditions: "a distinctive behaviour pattern shared by two or more individuals in a social unit, which persists over time and that new practitioners acquire in part through socially aided learning"¹⁷. Accordingly, for those happy to treat culture as a synonym for tradition, a phrase such as 'cultural evolution in chaffinch song' will sit unproblematically alongside hundreds of others identifying behavioural traditions in a variety of vertebrate (and perhaps invertebrate) species. For others, human culture involves so much that cannot be reduced to the existence of a tradition that they prefer to define culture more restrictively, at one extreme applying criteria such as language and symbolism, which limit culture to humans alone: animals have traditions—humans have culture.

Of more interest from an evolutionary perspective is a focus on characteristics that are shared to a greater or lesser extent with other species, including particular transmission mechanisms such as imitation or teaching, which apply to a smaller set of species than those shown to display traditions^{24,25,37}. When tradition and culture are distinguished in this way, it is common to regard tradition as the more basic and widespread phenomenon that, in certain restricted contexts, has evolved into more refined forms worth distinguishing as culture. This offers more scope for comparative analysis than simply equating tradition and culture, but can leave us with a simplistic debate over which species do or do not have culture. As the available data become richer, a more productive approach might be to dissect culture into multiple elements for comparative analysis^{15,38}, accepting that the resulting picture might indicate mosaic-like evolutionary patterns rather than a unitary pathway of cultural elaboration. This review outlines a compressed analysis of this kind, focusing on three aspects of cultural complexity.

Box 2 | The different social conventions of neighbours

The 'grooming hand-clasp' was the first social custom to be identified in chimpanzees, routine at Mahale but absent at Gombe, just 100 km away. Recently, it was discovered that although the Mahale K community (photographed, but now extinct) used the originally described palm-to-palm convention (left), members of the neighbouring M community typically show a different, wrist-to-wrist hand-clasp (right)³⁹. Moreover, the relative status of the groomers is apparent in the placement of the hands. Gwekulo, an adult female that transferred from the K to the M community, was observed to adopt the preferred wrist-to-wrist pattern of her new partners some of the time, but also to influence them to occasionally make palm-to-palm contact; however, she made delicate adjustments to do so, flexing her elbow in the local customary way, rather than keeping it straight, which was the norm in K community⁴⁰.



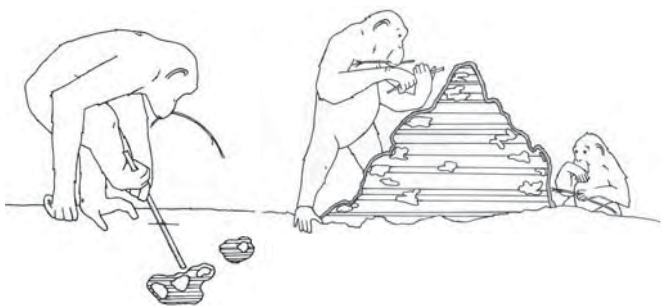
Box 2 Figure | Hand-clasp styles. Left, palm-to-palm (drawing by D. Bygott). Right, wrist-to-wrist (courtesy of M. Nakamura).

Constraints on such experimentation in the wild, such as the addition of conditions to control for individual learning in cases like the spread of *Coula*-cracking, has meant that social learning has been studied more extensively in captivity. However, the first such experiment seeding an expert tool-user into a naive community found surprisingly little evidence of imitation²⁸. After a flurry of similar findings, doubters soon began to ask: “Do apes ape?”²⁹ after all. A raft of more focused experiments ensued, and a recent review of ape studies over the past 15 years (ref. 30) concluded that, perhaps as a consequence of greater diversity in methodological approaches, the pendulum has swung back: out of 31 experiments, 22 involved chimpanzees and of these, 10 reported imitative behaviour and another 5 reported ‘emulation’, in which learning is focused on the outcomes of what the model achieved rather than precisely how it was done. The emerging picture is that apes do ape, but that imitation is just one of a ‘portfolio’ of varied social learning processes³¹, and perhaps most interestingly of all, it is applied selectively (Box 4).

This selectivity is one probable explanation for the curious mixture of both impressive and negative findings that now co-exist in the burgeoning literature on social learning in apes³¹. Among the impressive qualities are the ability to recognize what it means to be asked to imitate a new action and to do so with significant success (‘do-as-I-do’), and to recognize and test that one’s own actions are being imitated by others³⁰—abilities that primates other than apes have failed to demonstrate. This echoes a similar and well-established difference between primates in the ability to use one’s image in a mirror⁵. This correlation may reflect a fundamental aspect of ape cognition that underlies the translation of perceived actions into one’s own actions³¹. Turning to the imitation of functional actions such as foraging techniques, experiments have demonstrated the copying of alternative sequential organizations applied to a series of actions used to open up complex ‘artificial fruits’³¹.

Box 3 | A tool-set for harvesting termites

At many study sites, chimpanzees harvest termites using a single probing tool inserted into the sides of the insects’ mounds. This skill has recently been shown to be acquired much earlier in young females, which spend more time than males closely observing the proficient fishing of their mothers⁴¹. Such evidence suggests that this skill is acquired by social learning. Recently, a study in the Goulougo Triangle, Republic of Congo, described chimpanzees approaching termite mounds already armed with appropriate tools, sometimes two different ones²³. The first is a stout stick (left), which is thrust into the ground using both hands and often a foot, puncturing a tunnel into the nest about 30 cm beneath the ground. A more delicate probe is then inserted into the tunnel to extract termites; this probe is first prepared by biting it to length, manually stripping the leaves and pulling it through the teeth to create an effective ‘brush-tip’. This brush-tip method, like the use of the puncturing stick, is not known for chimpanzees harvesting termites elsewhere in Africa. The drawing on the right shows a female ready with such a probe in her mouth, and holding a third tool-type used for perforating termite mounds. Images drawn by D. Morgan from a video by C. Sanz and D. Morgan.



Nevertheless, when chimpanzees and children are compared on similar versions of the same task, children typically show higher copying fidelity. The findings summarized in Box 4 illustrate a common difference found in such studies. The observational learning of chimpanzees is highly pragmatic, subjugated to individual efforts wherever this gets results. In contrast, children are more prone to copy the actions of others just because others are doing them, betraying an extreme form of reliance on cultural convention. This difference was highlighted in a recent study in which chimpanzees and children observed conspecific models struggling to open two parts of an artificial food item: the children later targeted the part struggled with, perhaps seeing this as the one the model aimed to open, whereas chimpanzees tended to choose the other part, pragmatically avoiding the one already shown to be problematic³².

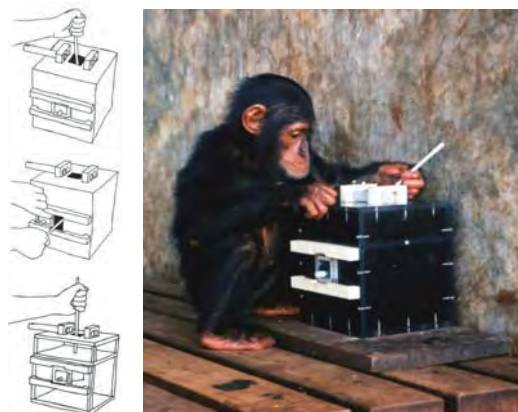
The contents of culture

Whatever the similarities and differences between chimpanzees, humans and other taxa in terms of the population-level patterning of traditions and underlying transmission mechanisms, a further set of comparative questions exists concerning the contents of those traditions. Again, there are significant overlaps as well as profound differences between human and chimpanzee behaviour. An example of overlap is the use of a wide range of types of tools; for example, chimpanzees use natural hammers, anvils, probes, sponges, clubs and seats. Key differences between the species in this domain include the construction of tools from complementary parts (as when hunter-gatherers construct even the simplest animal traps) and the use of weapons for hunting, both of which are found in humans but not chimpanzees. Chimpanzees fashion tools, but only by removing and modifying subcomponents¹⁵ (see Box 3).

Other overlaps and differences beyond the scope of this piece can be dissected into such domains as foraging, grooming, courtship and social behaviour¹⁵. An interesting example that has now been

Box 4 | Selective copying: are chimpanzees more rational imitators than children?

In a recent experiment⁴², young children and chimpanzees watched a familiar human first stab a tool into a small hole in the top of an opaque box (top diagram), then remove it and apply it to a second, lower hole to recover food (middle diagram). When allowed their turn (photo), youngsters in both groups typically tackled both holes in the sequence demonstrated. In a second condition, the box was transparent and revealed that the first stabbing action from the top was in fact ineffectual (bottom diagram). Now, the chimpanzees switched from the imitative approach they had used with the opaque box to a relatively emulative strategy that focused on the crucial terminal work around the lower hole. In contrast, the children continued to ‘blindly’ imitate the original stabbing sequence, which although apparently less rational in this particular context, emphasizes the extremes of conformity to which our own, super-cultural species is often subject.



comprehensively analysed is self-medication, which involves swallowing whole, rough-surfaced leaves (chimpanzees only) and chewing bitter piths (both chimpanzees and humans)³³. Extensive correlational evidence suggests that these actions counter intestinal worm infections, and recent additional experimental evidence³³ implies that innate predispositions in chimpanzees to indulge in such behaviour are refined by observational learning—an adaptive package given the requirement in the wild to discriminate specific local items from highly poisonous alternatives. Both the function of such self-medication and its social mode of acquisition appear to be common to chimpanzees and humans, although the human repertoire is more extensive.

Conclusions, controversies and the future

The emerging picture outlined here is significant for both chimpanzees and humans. With regards to chimpanzees, we now know that not only are chimpanzees as a species endangered, but unique local cultures are being destroyed at an even faster pace, much like some contemporary small-scale human cultures. A more positive effect of this research is that as increasing links between chimpanzee and human behaviour are discovered, people may become more motivated in their conservation efforts. Among the many implications for humans is a deeper understanding of both the shared and unique mental heritage that underlies our cultural capacity. A challenge for comparative genomics will be to help explain these similarities and differences. Of particular interest is the sheer scale of chimpanzee culture. Although all animals that acquire traditions may benefit from this second kind of inheritance system, the richness of chimpanzee culture outlined here suggests that this system may be particularly significant in this species and might interact with processes of genetic inheritance in interesting ways.

The synthesis of recent work on this topic has generated much new knowledge, along with a number of controversies that will drive future research. One of the fundamental controversies concerns the explanation for the evolution of distinctive cumulative culture in humans. A dominant view has been that deficits in the social learning capacity of chimpanzees, coupled with other aspects of social cognition (see reviews 34, 35), account for this difference^{24,25}. However, if chimpanzee social learning is as sophisticated as some of the recent work reviewed above suggests, this view becomes less tenable¹⁵. An alternative hypothesis is that the important differences lie in the cognitive complexity of the relevant cultural contents: the capacity to knap an Acheulian biface, for example, as much as the capacity to copy the skills of others¹⁵. Experimental programmes are now underway that may help to resolve this controversy by tracking the cultural dynamics of whole groups of chimpanzees³⁶, joining full-circle with the knowledge that continues to be garnered from the African forests.

- Goodall, J. Tool-using and aimed throwing in a community of free-living chimpanzees. *Nature* **201**, 1264–1266 (1964).
- Goodall, J. *The Chimpanzees of Gombe: Patterns of Behavior* (Harvard Univ. Press, Cambridge, 1986).
- Wrangham, R. W., McGrew, W. C., de Waal, F. B. M. & Heltne, P. (eds) *Chimpanzee Cultures* (Harvard Univ. Press, Cambridge, 1994).
- Nishida, T., Kano, T., Goodall, J., McGrew, W. C. & Nakamura, M. Ethogram and ethnography of Mahale chimpanzees. *Anthropol. Sci.* **107**, 141–188 (1999).
- Suddendorf, T. & Whiten, A. Mental evolution and development: Evidence for secondary representation in children, great apes and other animals. *Psychol. Bull.* **127**, 629–650 (2001).
- Boesch, C. & Hohmann, G. & Marchant, L. F. (eds) *Behavioural Diversity in Chimpanzees and Bonobos* (Cambridge Univ. Press, Cambridge, 2002).
- Matsuzawa, T. (ed.) *Chimpanzees of Bossou and Nimba 1976–2001* (Primate Research Institute, Kyoto Univ., Kyoto, 2002).
- McGrew, W. C. *The Cultured Chimpanzee: Reflections on Cultural Primatology* (Cambridge Univ. Press, Cambridge, 2004).
- Reynolds, V. *The Chimpanzees of the Budongo Forest: Ecology, Behaviour and Conservation* (Oxford Univ. Press, Oxford, 2005).
- The Chimpanzee Sequencing and Analysis Consortium. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* doi:10.1038/nature04072 (this issue).

- Caldecott, J. & Miles, L. (eds) *The World Atlas of Great Apes and their Conservation* (UNEP World Conservation Monitoring Centre, Univ. California Press, Berkeley, 2005).
- Whiten, A. et al. Cultures in chimpanzees. *Nature* **399**, 682–685 (1999).
- Whiten, A. et al. Charting cultural variation in chimpanzees. *Behaviour* **138**, 1489–1525 (2001).
- van Schaik, C. P. et al. Orangutan cultures and the evolution of material culture. *Science* **299**, 102–105 (2003).
- Whiten, A., Horner, V. & Marshall-Pescini, S. Cultural panthropology. *Evol. Anthropol.* **12**, 92–105 (2003).
- Lefebvre, L. in *The Evolution of Cognition* (eds Heyes, C. & Huber, L.) 311–328 (MIT Press, Cambridge, 2000).
- Fragaszy, D. M. & Perry, S. (eds) *The Biology of Traditions: Models and Evidence* (Cambridge Univ. Press, Cambridge, 2003).
- McGrew, W. C. *Chimpanzee Material Culture: Implications for Human Evolution* (Cambridge Univ. Press, Cambridge, 1992).
- Rendell, L. & Whitehead, H. Cultures in whales and dolphins. *Behav. Brain Sci.* **24**, 309–382 (2001).
- Perry, S., et al. in *The Biology of Traditions: Models and Evidence* (eds Fragaszy, D. M. & Perry, S.) 391–425 (Cambridge Univ. Press, Cambridge, 2003).
- Boesch, C. Is culture a golden barrier between humans and chimpanzees? *Evol. Anthropol.* **12**, 82–91 (2003).
- Nishida, T., Mitani, J. & Watts, D. Variable grooming behaviours in wild chimpanzees. *Folia Primatol. (Basel)* **75**, 31–36 (2004).
- Sanz, C., Morgan, D. & Glick, S. New insights into chimpanzees, tools and termites from the Congo Basin. *Am. Nat.* **164**, 567–581 (2004).
- Richerson, P. J. & Boyd, R. *Not by Genes Alone: How Culture Transformed Human Evolution* (Chicago Univ. Press, Chicago, 2005).
- Tomasello, M. *The Cultural Origins of Human Cognition* (Harvard Univ. Press, Cambridge, 1999).
- Matsuzawa, T. et al. in *Primate Origins of Human Cognition and Behavior* (ed. Matsuzawa, T.) 557–574 (Springer, Berlin, 2001).
- Biro, D. et al. Cultural innovation and transmission of tool use in wild chimpanzees: evidence from field experiments. *Anim. Cogn.* **6**, 213–223 (2004).
- Tomasello, M., Davis-Dasilva, M., Camak, L. & Bard, K. Observational learning of tool use by young chimpanzees and enculturated chimpanzees. *Hum. Evol.* **2**, 175–183 (1987).
- Tomasello, M. in *Social Learning in Animals: the Roots of Culture* (eds Heyes, C. M. & Galef, B. G.) 319–346 (Academic, London, 1996).
- Nielsen, M., Collier-Baker, E., Davis, J. M. & Suddendorf, T. Imitation recognition in a captive chimpanzee (*Pan troglodytes*). *Anim. Cogn.* **8**, 31–36 (2005).
- Whiten, A., Horner, V., Litchfield, C. A. & Marshall-Pescini, S. How do apes ape? *Learn. Behav.* **32**, 36–52 (2004).
- Call, J., Carpenter, M. & Tomasello, M. Copying results and copying actions in the process of social learning: chimpanzees (*Pan troglodytes*) and human children (*Homo sapiens*). *Anim. Cogn.* **8**, 151–163 (2005).
- Huffman, M. A. & Hirata, S. An experimental study of leaf swallowing in captive chimpanzees: insights into the origin of a self-medicative behaviour and the role of social learning. *Primates* **45**, 113–118 (2004).
- Hauser, M. Our chimpanzee mind. *Nature* doi:10.1038/nature03917 (this issue).
- de Waal, F. B. M. A century of getting to know the chimpanzee. *Nature* doi:10.1038/nature03999 (this issue).
- Whiten, A., Horner, V. & de Waal, F. B. M. Conformity to cultural norms of tool use in chimpanzees. *Nature* advance online publication, 21 August 2005 (doi:10.1038/nature04047).
- Galef, B. G. Jr. The question of animal culture. *Hum. Nat.* **3**, 157–178 (1992).
- Byrne, R. W. et al. Understanding culture across species. *Trends Cog. Sci.* **8**, 341–346 (2005).
- McGrew, W. C., Marchant, L. F., Scott, S. E. & Tutin, C. E. G. Intergroup differences in a social custom in wild chimpanzees: the grooming handclasp of the Mahale Mountains. *Curr. Anthropol.* **42**, 148–153 (2001).
- Nakamura, M. & Uehara, S. Proximate factors of different types of grooming hand-clasp in Mahale chimpanzees: implications for chimpanzee social customs. *Curr. Anthropol.* **45**, 108–114 (2004).
- Lonsdorf, E. V., Pusey, A. E. & Eberly, L. Sex differences in learning in chimpanzees. *Nature* **428**, 715–716 (2004).
- Horner, V. & Whiten, A. Causal knowledge and imitation/emulation switching in chimpanzees (*Pan troglodytes*) and children. *Anim. Cogn.* **8**, 164–181 (2005).

Acknowledgements The author was supported by a Leverhulme Major Research Fellowship. I am grateful to V. Horner, W. C. McGrew, D. Morgan and M. Nakamura for comments on the manuscript, and to S. Smart and J. Allen for image processing.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The author declares no competing financial interests. Correspondence and requests for materials should be addressed to A.W. (a.whiten@st-and.ac.uk).

PROGRESS

A century of getting to know the chimpanzee

Frans B. M. de Waal¹

A century of research on chimpanzees, both in their natural habitat and in captivity, has brought these apes socially, emotionally and mentally much closer to us. Parallels and homologues between chimpanzee and human behaviour range from tool-technology and cultural learning to power politics and intercommunity warfare. Few behavioural domains have remained untouched by this increased knowledge, which has dramatically challenged the way we view ourselves. The sequencing of the chimpanzee genome will no doubt bring more surprises and insights. Humans do occupy a special place among the primates, but this place increasingly has to be defined against a backdrop of substantial similarity.

As long ago as Plato's failed attempt to define Man as the only creature at once featherless and walking on two legs (in response to which Diogenes arrived in the lecture hall with a plucked chicken), humanity has been hard-pressed to find ultimate proof of its uniqueness. The manufacture of tools, for example, was once regarded as so special that a book appeared under the title *Man the Tool-Maker*¹. This definition held until wild chimpanzees (*Pan troglodytes*) were discovered modifying twigs to make them suitable for termite fishing. Another claim of human uniqueness concerned language, initially defined as symbolic communication. As soon as linguists heard about apes that had learned American Sign Language, however, they replaced the symbol requirement with their current emphasis on syntax. Humanity's special place in the cosmos is one of abandoned claims and moving goalposts.

The more we learn about apes, the more they seem as similar to us as their genetic material implies. Study of their behaviour began early last century with a handful of laboratory scientists. Wolfgang Köhler described how chimpanzees faced with an out-of-reach banana in the presence of boxes and sticks would sit around until the solution suddenly struck them: a flash of insight still referred to by insiders as a 'Köhler-moment'². Robert Yerkes documented the temperament of apes, and conducted pioneering experiments on cognition and cooperation³. Nadezhda Ladygina-Kohts followed in Charles Darwin's footsteps by offering a point-by-point comparison of the emotional expressions of a young chimpanzee and a human child⁴ (Fig. 1).

In those days, work in the natural habitat was frowned upon as unscientific: only laboratory approaches provided the controls required for conclusive science. Tension between these approaches persists today, even though the history of chimpanzee research is a showcase for the power of cross-fertilization between laboratory and field. The next series of insights came from attempts to study wild chimpanzees. At first, these attempts consisted of brief excursions, such as Henry Nissen's three-month stay in Guinea in the 1930s for the purpose of documenting chimpanzee feeding habits⁵. It was only in the 1960s that two pioneering long-term projects were initiated, and these were to inspire many more. On the Eastern shore of Lake Tanganyika in Tanzania, Jane Goodall set up camp in the Gombe Stream Reserve, and Toshisada Nishida did the same 170 km to the south, in the Mahale Mountains.

Studies in the field shattered the image of chimpanzees as peaceful vegetarians and began to reveal their astonishing social complexity. Meat consumption had been considered uniquely human among the primates, but chimpanzees were observed to catch colobus monkeys

(*Colobus badius*), tear them apart and eat them alive⁶. Although the initial impression of chimpanzees had been that they lack social bonds (except for the tie between mothers and dependent offspring), it was discovered that all individuals in a particular stretch of forest meet regularly. However, interactions with individuals in neighbouring areas, if they occur at all, tend to be negative⁷. Field workers began to speak of 'communities' in order to avoid the term 'group', as



Figure 1 | One of the first cognitive primatologists: Nadia Kohts. From 1913–1916, Nadezhda Ladygina-Kohts (also known as Nadia Kohts) raised a young chimpanzee, Joni, in her Moscow home⁴. She conducted tool, mirror, art and discrimination tasks, in the process inventing the still-popular matching-to-sample paradigm. Kohts reported a wide range of emotional responses in Joni, from jealousy and guilt to empathy and fierce loyalty to loved ones. She described Joni's facial expressions in muscle-by-muscle detail. Even though her cognitive and socio-emotional approach was far ahead of its time, Kohts is less well-known than some of her contemporaries, perhaps because of her gender and publication in Russian. Photograph taken by A. F. Kohts in 1914, and reproduced with permission from the State Darwin Museum in Moscow, Russia.

¹Living Links, Yerkes National Primate Research Center, Emory University, 954 North Gatewood Road, Atlanta, Georgia 30322, USA.



Figure 2 | Chimpanzees invite reconciliation by means of eye contact and hand gestures¹⁹. This scene occurred ten minutes after a protracted, noisy conflict between two adult males at a zoo. The challenged male (left) fled into the tree. He is now being approached by his opponent, who offers him an open hand. Within seconds, the two males had a physical reunion, kissed and embraced, then climbed to the ground to groom each other. Such peacemaking serves to maintain valuable relationships despite occasional conflict. Photograph by F. B. M. de Waal.

chimpanzees are rarely seen in large aggregations—they split up in ever-changing small ‘parties’ that travel through the forest, a system known as fission–fusion. Another claim of human uniqueness was abandoned when it was discovered that we are not the only primates to kill our own kind. Reports of lethal fighting between chimpanzee communities over territory⁸ profoundly affected the post-war debate about the origins of human aggression.

A second wave of influential chimpanzee studies in captivity in the 1970s placed them cognitively closer to humans than anyone had imagined. Gordon Gallup showed that apes recognize themselves in a mirror, indicating a level of self-awareness that sets humans and apes

apart from all other primates⁹. Emil Menzel conducted experiments in which an ape that knew where an item was hidden was released together with fellow apes that lacked such knowledge, and recorded how they learned from or outwitted one another¹⁰. This work set the stage for the ‘guesser’ versus ‘knower’ paradigm of modern inter-subjectivity research on apes and children¹¹. At about the same time, one of the world’s largest colonies of outdoor-living chimpanzees was established at the Arnhem Zoo in the Netherlands, where I documented machiavellian power politics and conflict resolution capacities¹².

Since then, field studies have continued to work on elucidating chimpanzee social organization¹³. Instead of considering the behaviour of this species a unitary phenomenon, there is increasing focus on behavioural and ‘cultural’ diversity from site to site¹⁴. This focus must be complemented by attention to genetic diversity, which will no doubt be stimulated by publication of the chimpanzee genome. Below, I will highlight three further areas of interest in chimpanzee behaviour: (1) aggression and conflict resolution, (2) reproductive strategies and (3) cooperation.

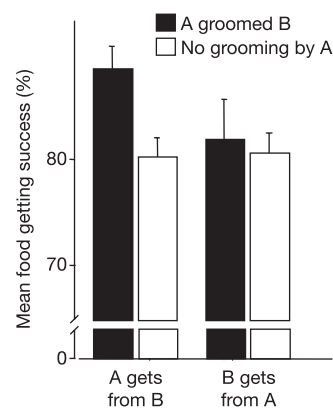
War and peace

Chimpanzee aggressive behaviour is quite different within and between groups. As with humans, intergroup aggression knows few inhibitions. A small group of chimpanzee males may stealthily enter a neighbouring territory to overwhelm a single enemy male that they viciously beat, bite and leave to die^{15,16}. Such attacks have actually been witnessed at a few field sites, whereas at other sites they have been strongly indicated. Initially, skeptics attributed chimpanzee ‘warfare’ to competition over the food that researchers provided in order to draw the apes out of the forest, but we now also have observations from unprovisioned sites.

Even though intragroup aggression occasionally turns deadly^{17,18}, it is far more constrained than intergroup aggression, owing to the adaptive value of group life. Intragroup tensions are actively coped with. After an open conflict, it is not uncommon for combatants to reunite with a kiss and embrace (Fig. 2). Known as reconciliation, this mechanism was discovered in captive chimpanzees¹⁹, has since been confirmed in the wild²⁰, and is in fact widespread in the primate



Figure 3 | Chimpanzees are known to reciprocally exchange favours. a, A cluster of chimpanzees gathers around branches with leaves provided in an experiment on reciprocity. The apes overcome their competitive tendencies and share the food. Photograph by F. B. M. de Waal. **b,** Mean ($\pm$ s.e.m.) success at getting food per dyadic direction between adult chimpanzees during food trials. Two conditions are distinguished: either individual A groomed B in the hours before branches with leaves were provided, or no



previous grooming of B by A occurred. The left-hand side of the graph shows the percentage of approaches by A to B that result in sharing by B (A gets from B); the right-hand side shows the same measure for B in obtaining food from A (B gets from A). A’s success in obtaining food from B was significantly higher after A had groomed B, whereas B’s success in obtaining food from A was unaffected by A’s previous grooming, strongly suggesting exchange of food for grooming³².

order and beyond²¹. The best-supported theory is that reconciliation serves to preserve valuable relationships despite the undermining effects of occasional conflict. Male chimpanzees are the more aggressive sex, but they are also the more conciliatory, which makes sense given that males stand more to lose if relationships deteriorate. They are together more often than females, and cooperate in hunting, intragroup politics and intergroup territoriality.

The deadliest form of intragroup aggression is aimed at young infants. Male infanticide is evolutionarily explained as the elimination of offspring sired by rivals, resulting in a shortened waiting time until a female's next ovulation²². This explanation proves problematic in relation to chimpanzees, however, as males sometimes kill infants they may well have fathered themselves^{23,24}. Infanticide by females is much rarer, and is thought to relate to food competition between females and their dependent offspring²⁵.

Sexual competition

Although humans have nuclear families, bonobos (*Pan paniscus*) and chimpanzees have none—our closest relatives are thoroughly promiscuous. Chimpanzee females mate with many different males, and bonobos have an even wider-ranging sex life, with frequent same-sex partners²⁶. Males are hardly involved in care for the young, and in fact often pose a threat (see above). It is thought that females mate with so many males partly to confuse the issue of paternity, thus countering male infanticide by making it hard for any male to exempt his own offspring²².

This promiscuous mating system explains the intense sexual rivalry among males as well as the size of their testes. Corrected for body size, chimpanzee testes are about 10 times larger than those of our own species²⁷. As females have multiple sex partners, sperm competition is likely: the higher the number of sperm cells per ejaculate, the better a male's chance of fertilization. No hominoid (that is, the primate family that includes humans and apes) apart from humans combines relatively small testes and minor sexual dimorphism. This suggests that human evolution has placed strong curbs on sexual competition, which may have been achieved by making mate choice less open-ended. Pair-bonding associated with male parental care probably traces as far back as *Australopithecus*²⁸.

The analysis of DNA from hair samples or fecal droppings is having a profound effect on our understanding of chimpanzee social structure. Genetic evidence can be used to determine whether male chimpanzees, which stay life-long in their natal community, are more closely related to each other than females, which tend to leave and join a neighbouring community around puberty. DNA data can also be used to determine paternity, so as to better understand what mating strategies actually lead to conception. Apart from confirming that the overwhelming majority of offspring are produced by intragroup fertilizations, thus explaining the observed male rivalry, it is too early to tell what these studies will reveal about chimpanzee social organization^{29,30}.

Quid pro quo

Chimpanzee society combines high levels of competition and cooperation. Cooperation is typical among kin (for example, mother and offspring) and among adult males, regardless of kinship. In pursuit of high status, young adult males operate mainly on the basis of fighting ability, but often cannot succeed without the support of older males. Like elder statesmen, post-prime males exert influence as alliance partners, without a chance of assuming top status themselves^{12,31}.

Political coalitions were recognized early on as part of an elaborate 'marketplace of services' in which chimpanzees trade grooming, sex, food and support¹². The rules of reciprocity governing social exchange are only beginning to be understood, but evidence is accumulating that chimpanzees repay both positive acts (for example, sharing food preferentially with previous grooming

partners³²; see Fig. 3) and negative acts (for example, squaring accounts with those who previously opposed them)³³. These tendencies are known in humans as 'gratitude' and 'retribution', respectively.

Perhaps the highest levels of cooperation and reciprocity have been observed during hunting. The chimpanzee diet includes substantial amounts of vertebrate meat³⁴. The hunt of colobus monkeys in some locations is so difficult that hunting skills take years to develop, and pursuing males are said to adopt a role division (that is, adopting roles of driver, blocker and ambusher). As in humans, the oldest males tend to take on the most difficult hunting tasks³⁵. The division of meat is a process of begging and sharing in which power, sex, and *quid pro quo* seem to meet in ways that are not yet fully understood. In one wild community, an alpha male with a tenure of more than a decade was described as having a 'bribery' system: he selectively shared prized meat with those that supported him or those from which support proved useful in the future³⁶. To unravel these and other complexities is a daunting task for the relatively small number of devoted scientists who continue to work on chimpanzee behaviour in both captivity and the wild.

- Oakley, K. *Man the Tool-Maker* (Univ. of Chicago Press, Chicago, 1957).
- Köhler, W. *Mentality of Apes* (Routledge & Kegan, London, 1925).
- Yerkes, R. M. *Chimpanzees: A Laboratory Colony* (Yale Univ. Press, New Haven, 1943).
- Ladygina-Kohts, N. N. in *Infant Chimpanzee and Human Child: A Classic 1935 Comparative Study of Ape Emotions and Intelligence* (ed. de Waal, F. B. M.) (Oxford Univ. Press, Oxford, 2001).
- Nissen, H. A field study of the chimpanzee. *Comp. Psychol. Monogr.* **8**, 1–122 (1931).
- Goodall, J. in *Primate Behavior: Field Studies of Monkeys and Apes* (ed. DeVore, I.) 425–473 (Holt, Rinehart & Winston, New York, 1965).
- Nishida, T. The social group of wild chimpanzees in the Mahali Mountains. *Primates* **9**, 167–224 (1968).
- Goodall, J. et al. in *The Great Apes* (eds Hamburg, D. A. & McCown, E. R.) 13–53 (Benjamin/Cummings, Menlo Park, California, 1979).
- Gallup, G. G. Jr. Chimpanzees: Self-recognition. *Science* **167**, 86–87 (1970).
- Menzel, E. W. in *Behavior of Non-human Primates* Vol. 5 (eds Schrier, A. M. & Stollnitz, F.) 83–153 (Academic, New York, 1974).
- Hauser, M. Our chimpanzee mind. *Nature* doi:10.1038/nature03917 (this issue).
- de Waal, F. B. M. *Chimpanzee Politics: Power and Sex Among Apes* (Jonathan Cape, London, 1982).
- Mitani, J. C., Watts, D. P. & Muller, M. Recent developments in the study of wild chimpanzee behaviour. *Evol. Anthropol.* **11**, 9–25 (2002).
- Whiten, A. The second inheritance system of chimpanzees and humans. *Nature* doi:10.1038/nature04023 (this issue).
- Goodall, J. *The Chimpanzees of Gombe* (Harvard Univ. Press, Cambridge, 1986).
- Wilson, M. L. & Wrangham, R. W. Intergroup relations in chimpanzees. *Annu. Rev. Anthropol.* **32**, 363–392 (2003).
- de Waal, F. B. M. The brutal elimination of a rival among captive male chimpanzees. *Ethol. Sociobiol.* **7**, 237–251 (1986).
- Watts, D. P. Intragroup coalitionary killing of an adult male chimpanzee at Ngogo, Kibale National Park, Uganda. *Int. J. Primatol.* **25**, 507–521 (2004).
- de Waal, F. B. M. & van Roosmalen, A. Reconciliation and consolation among chimpanzees. *Behav. Ecol. Sociobiol.* **5**, 55–66 (1979).
- Wittig, R. M. & Boesch, C. "Decision-making" in conflicts of wild chimpanzees (*Pan troglodytes*): An extension of the Relational Model. *Behav. Ecol. Sociobiol.* **54**, 491–504 (2003).
- Aureli, F. & de Waal, F. B. M. *Natural Conflict Resolution* (Univ. California Press, Berkeley, 2000).
- Hrdy, S. B. Infanticide among animals: A review, classification, and examination of the implications for the reproductive strategies of females. *Ethol. Sociobiol.* **1**, 13–40 (1979).
- Hiraiwa-Hasegawa, M. & Hasegawa, T. in *Infanticide and Parental Care* (eds Parmigiani, S. & vom Saal, F. S.) 137–154 (Harwood Academy, London, 1994).
- Arcadi, A. & Wrangham, R. Infanticide in chimpanzees: Review of cases and a new within-group observation from the Kanyawara study group in Kibale National Park. *Primates* **40**, 337–351 (1999).
- Pusey, A. E., Williams, J. & Goodall, J. The influence of dominance rank on the reproductive success of female chimpanzees. *Science* **277**, 828–831 (1997).
- de Waal, F. B. M. Tension regulation and nonreproductive functions of sex among captive bonobos. *Nat. Geogr. Res.* **3**, 318–335 (1987).
- Harcourt, A. H., Harvey, P. H., Larson, S. G. & Short, R. V. Testis weight, body weight and breeding system in primates. *Nature* **293**, 55–57 (1981).
- Reno, P. L., Meindl, R. S., McCollum, M. A. & Lovejoy, C. O. Sexual dimorphism in *Australopithecus afarensis* was similar to that of modern humans. *Proc. Natl Acad. Sci. USA* **100**, 9404–9409 (2003).

29. Constable, J., Ashley, M., Goodall, J. & Pusey, A. Noninvasive paternity assignment in Gombe chimpanzees. *Mol. Ecol.* **10**, 1279–1300 (2001).
30. Vigilant, L., Hofreiter, M., Siedel, H. & Boesch, C. Paternity and relatedness in wild chimpanzee communities. *Proc. Natl Acad. Sci. USA* **98**, 12890–12895 (2001).
31. Nishida, T. Alpha status and agonistic alliance in wild chimpanzees. *Primates* **24**, 318–336 (1983).
32. de Waal, F. B. M. The chimpanzee's service economy: Food for grooming. *Evol. Hum. Behav.* **18**, 375–386 (1997).
33. de Waal, F. B. M. & Luttrell, L. M. Mechanisms of social reciprocity in three primate species: Symmetrical relationship characteristics or cognition? *Ethol. Sociobiol.* **9**, 101–118 (1988).
34. Stanford, C. B. *Chimpanzee and Red Colubus* (Harvard Univ. Press, Cambridge, 1998).
35. Boesch, C. in *Animal Social Complexity* (eds de Waal, F. B. M. & Tyack, P. L.) 93–110 (Harvard Univ. Press, Cambridge, 2003).
36. Nishida, T., Hasegawa, T., Hayaki, H., Takahata, Y., Uehara, S., et al. in *Topics in Primatology Vol. 1: Human Origins* (ed. Nishida, T.) 159–174 (Univ. Tokyo Press, Tokyo, 1992).

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The author declares no competing financial interests. Correspondence and requests for materials should be addressed to the author (dewaal@emory.edu).

PROGRESS

Our chimpanzee mind

Marc Hauser¹

Some might consider the title of this piece preposterous. Bishop Wilberforce would no doubt have shaken his fist at it, just as he disputed Huxley's championing of darwinian continuity. But the title of this essay is no more outrageous than one entitled 'The chimpanzee's bird brain', for there has been extensive evolutionary conservation of many neural and psychological functions across species. We share with chimpanzees some—but not all—mental functions, some of which are shared with other species as well. As the publication of the chimpanzee genome reveals, we also share a good deal of our DNA. Unfortunately, we are virtually in the dark when it comes to understanding how genes build minds. If comparative genomics is to enlighten our understanding of human origins, it must be accompanied by an equally rich description of animal psychology, both in terms of its underlying neural signatures and the evolutionary processes that led to convergence and divergence with other species.

Here I will focus on research in two domains of knowledge that have made considerable progress over the past five years: folk mathematics and folk psychology. What I mean by 'folk' is a sense of knowing that operates in the absence of education or other forms of culturally imposed experience—part of the brain's core knowledge^{1,2}. Although the focus here is on comparisons between chimpanzees and humans, much of what follows is unlikely to be specific to chimpanzees. In fact, the bottom line at present is that for each psychological capacity explored, some other animal shares this ability with chimpanzees. The reason why chimpanzees may be uniquely placed to enlighten human origins is due both to their phylogenetic proximity to humans as well as the extent to which they have accumulated a suite of psychological abilities in the service of solving social and ecological problems that were largely shared with those faced by our hominid hunter-gatherers^{3,4}.

Folk mathematics

It might seem bizarre to ask about the evolution of mathematics, a system that originated with *Homo sapiens*. But this achievement did not emerge *ex nihilo*. Human adults and infants with no tutoring have at least two core, folk mathematical systems^{1,5–7}. One operates by representing, in parallel, a small number (<5) of discrete objects or events. The second is unconstrained by magnitude, but its operation is approximate, constrained by the ratio between numbers. These two systems are not only present in chimpanzees, but evolved millions of years before the primates. What seems unique to humans is the integer list, a set of discrete symbols that enable us to quantify large numbers precisely.

A sense of number has a role in at least four naturally occurring contexts for chimpanzees and other animals: foraging, group hunting, food sharing and intercommunity warfare. Consider warfare. Along with humans, chimpanzees are among a handful of species willing to engage in lethal fighting (Fig. 1). Forty years of observations across Africa have shown that when three or more males from one community find a lone individual from a neighbouring community, they kill this individual. This ratio is meaningful, representing the minimum number of males necessary to hold and kill an intruder^{9,10}. Experiments¹¹ confirm this numerical ratio and power asymmetry: if the loud call of a foreign male is played over a loudspeaker, groups of fewer than three males remain silent and still, whereas groups of three or more males call back and move towards the speaker in preparation for an attack.

Like other animals, the capacity for numerical quantification plays an essential role in the socioecology of chimpanzees. Whatever we discover in the laboratory, therefore, is unlikely to represent an artefact of testing, even though laboratory studies might reveal additional or more refined abilities owing to methodological differences.

An early demonstration of numerical ability in animals—and the large approximate system in particular—used an operant procedure with rats and pigeons^{8,12}. After some number of events, such as light flashes or tones, animals depressed a key for food. The proportion of errors increased with the magnitude of the target number, with discrimination constrained by the ratio between numbers. A number discrimination problem with chimpanzees involved the sequential placement of individual food rewards into one of two concealed wells^{13,14}. Success consisted of picking the larger of the two food quantities. Performance was significantly above chance with ratios of food quantity below 0.70, independent of absolute number. Like other species, chimpanzees show the signature of the large approximate number system.

Initial evidence of the small precise system in animals emerged from studies of rhesus monkeys. Using looking-time as a measure of expectation¹⁵, animals looked for longer after a 1 + 1 operation if the outcome was 1 or 3 than if it was 2 (refs 16, 17). Subjects failed to show a difference in looking time when the outcome was greater than 3 or when there were multiple addition operations; for example, rhesus monkeys looked for equally long periods of time at an outcome of 3, 4 or 5 following a 1 + 1 + 1 operation¹⁸. These results reflect the small precise system, but not the large approximate system that is unconstrained by absolute number. Although studies of chimpanzees have yet to explore the small precise system, there is no reason to doubt its presence given evidence in other primates, including human infants.

Chimpanzee research departs from work on most other animals in training them to learn Arabic numerals^{19–23}. I focus here on Matsuzawa's studies of the chimpanzee Ai because they are the most extensive, documenting both the acquisition of a number system as well as its limitations. After several years, Ai learned the first nine integers, acquiring an understanding of ordinality and cardinality. For example, when presented with three-to-five Arabic numerals on a monitor, Ai pressed each number in its appropriate ordinal sequence, independent of numerical distance. Like humans, however, Ai's response was fastest with greater distances, and the number of errors

¹Departments of Psychology, Organismic and Evolutionary Biology, and Biological Anthropology, Harvard University, Cambridge, Massachusetts 02138, USA.

increased with smaller ratios—patterns that reveal the signature of the large approximate number system (Fig. 2).

Closer inspection of Ai's performance reveals a different number sense from our own. Ai's initial training required mapping '1' and '2' to the correct number of food rewards. When then tested on '3', Ai failed to generalize, applying it indiscriminately to arrays of two or three objects. The same thing happened with each successive number added to the list: extensive training, followed by a failure to generalize. Ai never learned the rule that each new Arabic numeral symbolized a new cardinal value. Human children do something different: their understanding of the first three integers emerges slowly, but once acquired (at about 3–3.5 years of age), children spontaneously generalize to the remaining set of numbers in the integer list^{1,6}. Ai learned the integer list by associating each symbol with a discrete quantity. Human children, in contrast, first acquire an arbitrary list (the words for counting) and then make an induction from a limited sample to generate an infinite list of numbers.

We share with chimpanzees and other animals two core systems of folk mathematics. However, we depart from all other animals in our capacity to represent large numbers precisely. What enables this capacity, and much more, is presently unclear, but there are

interesting possibilities on the horizon. For example, not all human cultures express the large precise system, even though their language is as expressive as any other natural language^{24–26}. In these cultures, the two core systems are operative, with number words mapped on to the first few integers and then the use of 'many' for higher values. This raises the question of how particular aspects of our language faculty have uniquely transformed particular aspects of our thoughts. One possibility is that a set of computational mechanisms recruited by the language faculty (for example, recursive operations) is tapped by our mathematical faculty, allowing both systems the power of open-ended expression^{27,28}.

Folk psychology

In parallel with the discussion of folk mathematics, we want to understand whether chimpanzees and other animals have a folk psychology, a system of knowing that enables individuals to infer what others believe, desire and want^{29,30}. In our own species, these abilities gradually emerge over the first few years of life, reaching a celebratory crescendo at the age of 4–5 years. At this point, children can use the direction of eye gaze to infer what others know and to manipulate others' beliefs by lying.

The general consensus until the year 2000 was that animals, including chimpanzees, lacked all components of our folk psychology^{31–33}. Povinelli's studies of chimpanzees provided the strongest support for this conclusion³⁴. In these studies, a chimpanzee entered a test room and for each condition, begged for food from one of two experimenters. One experimenter could see the begging chimpanzee and the other could not, using body position, blindfolds, buckets and other devices to alter visual perspective. The chimpanzee begged equally from both experimenters, and never learned to beg selectively from only the experimenter who could see it. Povinelli concluded that chimpanzees fail to use seeing to infer what others know.

Using a different experimental approach, Hare and colleagues challenged Povinelli's conclusions^{35,36}. Wild chimpanzees compete with each other more often than they cooperate. Povinelli's experiments involved cooperative communication between chimpanzee and human, but Hare's experiments involved competition between two chimpanzees of different dominance ranks, and were designed to test whether individuals use information about seeing to make inferences about knowing. Each condition imposed different

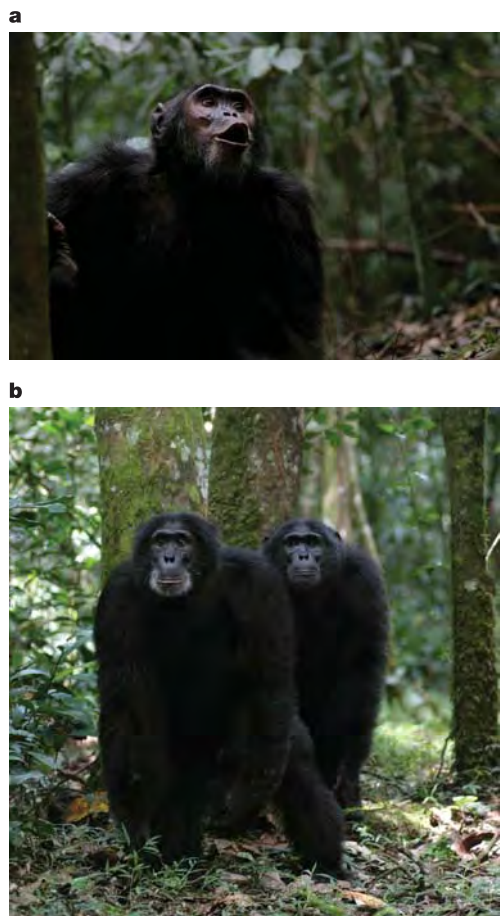


Figure 1 | Dangerous encounters. Chimpanzees compete with conspecifics both within and between communities. They may encounter members of other communities by hearing vocalizations of their neighbours (a), in which case they can become worried and their hair becomes erect (b). They can also see foreign chimpanzees during boundary patrols, when they make deep incursions into the territories of neighbouring communities. During such patrols, which are undertaken primarily by adult males, the chimpanzees are unusually silent and wary, and do not stop to feed. If they encounter chimpanzees from another community, fights may ensue and sometimes result in fatalities. Photographs courtesy of Kevin Langergraber at the University of Michigan, Ann Arbor, Michigan, USA.

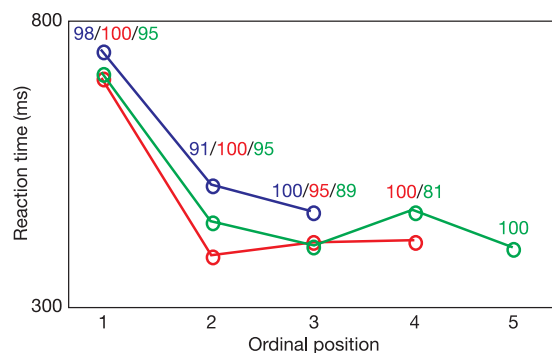


Figure 2 | The responses of chimpanzee Ai to a serial order task involving Arabic numerals. The y axis shows reaction times (in milliseconds) and the x axis shows the ordinal position of the number presented on the monitor. For example, '1' refers to the smallest number presented on the monitor, and the number that should have been pressed first in the ordinal sequence. If there were four numbers, '4' represents the largest number for the presented sequence. Blue symbols refer to three-item lists, red to four-item lists, and green to five-item lists. Above each cluster of points is the percentage of correct responses (that is, for contacting the correct number in the ordinal sequence, for three/four/five-item lists). Reaction time was slowest before the first selection, and then flattened out for the rest of the sequence, suggesting an initial planning stage. Graph generated from data presented in ref. 22.

constraints on what either one or both individuals could see. Consider three conditions in which the experimenter opened the door to the subordinate's room slightly before opening the dominant's door; this allowed the subordinate chimpanzee to make the first move in the absence of the dominant's behaviour (Fig. 3). Condition 1 involved one banana visible to both competitors, and one banana hidden behind an opaque barrier and visible only to the subordinate. Condition 2 involved placing one banana visible to both, and one behind a transparent barrier, so that both bananas were in view. Condition 3 involved two opaque barriers. While the subordinate chimpanzee watched, and the dominant chimpanzee looked away, the experimenter concealed one banana on the subordinate's side of the barrier. This condition differs from condition 1 in that it asks whether subordinates attend to the mere presence of a dominant or, more importantly, to what the dominant can see. If mere presence dictates competitive behaviour, then subordinates should stay put. In contrast, if subordinates recognize that the dominant chimpanzee failed to see the baiting of the food, then they should move out and pick up the concealed banana.

In condition 1, subordinates retrieved about half of the food, typically moving to the barrier before the dominant reacted. In condition 2, subordinates stayed put, allowing the dominant to run out and grab both bananas. The success of the subordinate chimpanzee in condition 1 was not attributable to physical protection from the barrier, as this would have worked equally well in condition 2. In condition 3, subordinates obtained more food than the dominants. Dominant chimpanzees could not see the hidden banana, which allowed subordinates to rush to the correct barrier. These and other results^{36–38} show that chimpanzees can infer what another chimpanzee knows on the basis of what it sees.

Hare's results led to more penetrating studies of chimpanzees and other species, emphasizing the importance of using methods that tap spontaneous abilities and are sensitive to species-typical environments^{37–41}. Results suggest that chimpanzees are equipped with some aspects of human folk psychology, but that they are not unique among animals. In fact, chimpanzees appear to be less adept than dogs at using the direction of eye gaze to infer the location of hidden food. Four questions direct current research. First, what is the nature

of the chimpanzees' knowledge of the mental states of other chimpanzees? Second, to what extent is this knowledge a specialization for the domain of competition? Third, in what ways might chimpanzees differ from other animals? And finally, what specific abilities evolved in humans to enable our distinctive folk psychology, with its unlimited capacity to represent others' beliefs about beliefs?

Conclusions

We share with chimpanzees and other animals core aspects of folk mathematics and psychology. This conclusion sets up three broad questions that we can now begin to address. First, given the evidence for homology at the behavioural level, to what extent are there homologies at the genetic and neural levels? At the genetic level, the publication of the chimpanzee genome will lead to increased capacity to pinpoint homologies. However, we are woefully ignorant about how genes build brains, and how the electrical activity of the brain builds thoughts and emotions. The situation is nonetheless more promising today than it was five years ago, owing to the convergence of three disciplines: comparative genomics, animal psychology and developmental neuropsychology. For example, we are now beginning to understand the genetics of Williams syndrome and autism, deficits that strike at our folk psychology. By looking at the constellation of genes that underlie these disorders, their presence or absence in chimpanzees and other species, and the nature of each group's psychological limitations, the gap between genomics and psychology is shrinking.

Second, given the ways in which human and chimpanzee knowledge diverge, what computations uniquely enable human abilities? Many researchers implicate language, but the answer is unsatisfying without specifying the computational details. What we want to understand is how the language faculty or some other property of the mind, such as our capacity for storing and recollecting memories, enabled us to uniquely develop sophisticated mathematical tools, teach, and to pass down rich cultures through stories and written histories.

Third, although the behavioural repertoire of the chimpanzee is more like our own than any other species^{3,4}, the psychological mechanisms underlying chimpanzee behaviour seem to be largely

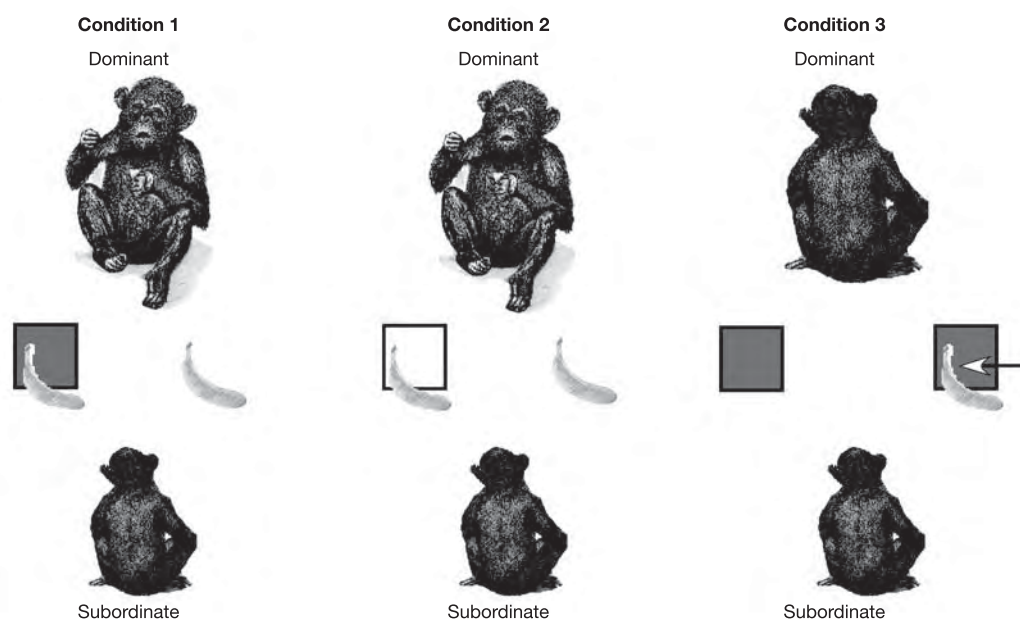


Figure 3 | Three conditions designed to test what chimpanzees know about seeing. In condition 1, one banana is visible to both the subordinate and dominant chimpanzee, but only the subordinate can see the banana behind the opaque screen. In condition 2, both the subordinate and dominant

chimpanzee can see both bananas, as the screen is transparent. In condition 3, one banana is introduced behind the opaque screen while the dominant chimpanzee looks away. Adapted from refs 35, 36.

shared with other species. That is, for every cognitive mechanism explored in chimpanzees (beyond those of folk mathematics and psychology), there are parallels in other species. This means either that chimpanzees use a suite of shared psychological capacities to solve different problems from other animals, or that our analyses are relatively superficial. My bet is on the latter, given that the history of neuroscience reveals a high degree of coupling between particular socioecological pressures and adaptive solutions. With the welcome addition of comparative genomics to comparative neurobiology and behaviour, answers to these questions are indeed on the horizon.

1. Hauser, M. D. & Spelke, E. S. in *The Cognitive Neurosciences* (ed. Gazzaniga, M.) 1025–1147 (MIT Press, Cambridge, 2004).
2. Spelke, E. Core knowledge. *Am. Psychol.* **55**, 1233–1243 (2000).
3. Whiten, A. The second inheritance system of chimpanzees and humans. *Nature* doi:10.1038/nature04023 (this issue).
4. de Waal, F. B. M. A century of getting to know the chimpanzee. *Nature* doi:10.1038/nature03999 (this issue).
5. Butterworth, B. *What Counts: How Every Brain is Hardwired for Math* (Free Press, New York, 1999).
6. Carey, S. Evolutionary and ontogenetic foundations of arithmetic. *Mind Lang.* **16**, 37–55 (2001).
7. Dehaene, S. *The Number Sense* (Oxford Univ. Press, Oxford, 1997).
8. Gallistel, C. R. *The Organization of Learning* (MIT Press, Cambridge, 1990).
9. Manson, J. H. & Wrangham, R. W. Intergroup aggression in chimpanzees and humans. *Curr. Anthropol.* **32**, 369–390 (1991).
10. Wrangham, R. W. & Peterson, D. *Demonic Males* (Houghton Mifflin, New York, 1996).
11. Wilson, M. L., Hauser, M. D. & Wrangham, R. W. Does participation in cooperative intergroup conflict depend on numerical assessment, range location, or rank for wild chimpanzees? *Anim. Behav.* **61**, 1201–1213 (2001).
12. Platt, J. R. & Johnson, D. M. Localization of position within a homogeneous behaviour chain: effects of error contingencies. *Learn. Motiv.* **2**, 386–414 (1971).
13. Beran, M. J. Summation and numerosness judgments of sequentially presented sets of items by chimpanzees (*Pan troglodytes*). *J. Comp. Psychol.* **115**, 181–191 (2001).
14. Beran, M. J. & Rumbaugh, D. M. "Constructive" enumeration by chimpanzees (*Pan troglodytes*) on a computerized task. *Anim. Cogn.* **4**, 81–89 (2001).
15. Wynn, K. Addition and subtraction by human infants. *Nature* **358**, 749–750 (1992).
16. Hauser, M. D., MacNeilage, P. & Ware, M. Numerical representations in primates. *Proc. Natl Acad. Sci. USA* **93**, 1514–1517 (1996).
17. Uller, C., Hauser, M. D. & Carey, S. Spontaneous representation of number in cotton-top tamarins (*Saguinus oedipus*). *J. Comp. Psychol.* **115**, 248–257 (2001).
18. Hauser, M. D. & Carey, S. Spontaneous number representations of small numbers of objects by rhesus macaques: examinations of content and format. *Cogn. Psychol.* **47**, 367–401 (2003).
19. Biro, D. & Matsuzawa, T. Numerical ordering in a chimpanzee (*Pan troglodytes*): planning, executing, and monitoring. *J. Comp. Psychol.* **113**, 178–195 (1999).
20. Boysen, S. T. & Bernston, G. G. Numerical competence in a chimpanzee. *J. Comp. Psychol.* **103**, 23–31 (1989).
21. Brannon, E. M. & Terrace, H. S. Ordering of the numerosities 1 to 9 by monkeys. *Science* **282**, 746–749 (1998).
22. Kawai, N. & Matsuzawa, T. Numerical memory span in a chimpanzee. *Nature* **403**, 39–40 (2000).
23. Tomonaga, M. & Matsuzawa, T. Sequential responding to arabic numerals with wild cards by the chimpanzee (*Pan troglodytes*). *Anim. Cogn.* **3**, 1–11 (2000).
24. Gelman, R. & Gallistel, C. R. Language and the origin of numerical concepts. *Science* **306**, 441–443 (2004).
25. Gordon, P. Numerical cognition without words: Evidence from Amazonia. *Science* **306**, 496–499 (2004).
26. Pica, P., Lemer, C., Izard, V. & Dehaene, S. Exact and approximate arithmetic in an Amazonian indigene group. *Science* **306**, 499–503 (2004).
27. Hauser, M. D., Chomsky, N. & Fitch, W. T. The faculty of language: What is it, who has it, and how did it evolve? *Science* **298**, 1569–1579 (2002).
28. Chomsky, N. *On Nature and Language* (Cambridge Univ. Press, New York, 2000).
29. Premack, D. & Premack, A. *Original Intelligence* (McGraw Hill, New York, 2002).
30. Premack, D. & Woodruff, G. Does the chimpanzee have a theory of mind? *Behav. Brain Sci.* **4**, 515–526 (1978).
31. Cheney, D. L. & Seyfarth, R. M. *How Monkeys See the World: Inside the Mind of Another Species* (Chicago Univ. Press, Chicago, 1990).
32. Hauser, M. D. *Wild Minds: What Animals Really Think* (Henry Holt, New York, 2000).
33. Tomasello, M. & Call, J. *Primate Cognition* (Oxford Univ. Press, Oxford, 1997).
34. Povinelli, D. J. & Eddy, T. J. What young chimpanzees know about seeing. *Monogr. Soc. Res. Child Dev.* **247**, 1–147 (1996).
35. Hare, B., Call, J., Agnetta, B. & Tomasello, M. Chimpanzees know what conspecifics do and do not see. *Anim. Behav.* **59**, 771–785 (2000).
36. Hare, B., Call, J. & Tomasello, M. Do chimpanzees know what conspecifics know? *Anim. Behav.* **61**, 139–151 (2001).
37. Call, J., Hare, B., Carpenter, M. & Tomasello, M. 'Unwilling' versus 'unable': chimpanzees' understanding of human intentional action. *Dev. Sci.* **7**, 488–498 (2004).
38. Hare, B. & Tomasello, M. Chimpanzees are more skilful in competitive than in cooperative cognitive tasks. *Anim. Behav.* **68**, 571–581 (2004).
39. Flombaum, J. & Santos, L. Rhesus monkeys attribute perceptions to others. *Curr. Biol.* **15**, 1–20 (2005).
40. Hare, B., Brown, M., Williamson, C. & Tomasello, M. The domestication of social cognition in dogs. *Science* **298**, 1636–1639 (2002).
41. Emery, N. J. & Clayton, N. S. Effects of experience and social context on prospective caching strategies by scrub jays. *Nature* **414**, 443–446 (2001).

Acknowledgements For support during the writing of this article, I wish to thank the McDonnell Foundation, the Guggenheim Foundation and a National Science Foundation ROLE grant.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The author declares no competing financial interests. Correspondence and requests for materials should be addressed to the author (mdh@wjh.harvard.edu).

PROGRESS

Molecular insights into human brain evolution

Robert Sean Hill¹ & Christopher A. Walsh¹

Rapidly advancing knowledge of genome structure and sequence enables new means for the analysis of specific DNA changes associated with the differences between the human brain and that of other mammals. Recent studies implicate evolutionary changes in messenger RNA and protein expression levels, as well as DNA changes that alter amino acid sequences. We can anticipate having a systematic catalogue of DNA changes in the lineage leading to humans, but an ongoing challenge will be relating these changes to the anatomical and functional differences between our brain and that of our ancient and more recent ancestors.

Santiago Ramon y Cajal, widely regarded as the founder of modern neuroscience, recognized as early as the turn of the twentieth century that the human brain was not just larger than that of our ancestors, but it differed in its circuitry as well. Over the course of the last century these differences have been extensively studied at a histological level, although specifying the exact changes that distinguish the human brain has been elusive.

“The opinion generally accepted at that time that the differences between the brain of [non-human] mammals (cat, dog, monkey, etc) and that of man are only quantitative, seemed to me unlikely and even a little offensive to the human dignity... My investigations showed that the functional superiority of the human brain is intimately bound up with the prodigious abundance and unusual wealth of forms of the so-called neurons with short axon.” (Ref. 1, translated by J. DeFelipe).

Comparative differences in brain structure

Understanding the genetic changes that distinguish our brain from that of our ancestors starts with defining the key structural and functional differences between the human brain and that of other primates. Our brain is roughly three times the size of the chimpanzee brain, our nearest living relative, from which we diverged 7–8 million years ago, and about twice the size of pre-human hominids living as recently as 2.5 million years ago². The increased size particularly affects the cerebral cortex, the largest brain structure and seat of most higher cognitive functions. The cortex is a multi-layered sheet that is smooth in rodents, but folded in mammals with larger cortices (Fig. 1), allowing more cortex to squeeze into the limited volume of the head.

The enlarged cortex of great apes reflects a longer period of neuronal formation during pre-natal development, so that each dividing progenitor cell undergoes more cell cycles before stopping cell division³. Cortical progenitors undergo 11 rounds of cell division in mice⁴, at least 28 in the macaque³, and probably far more in human. In addition to making a larger cortex, the longer period of neurogenesis adds novel neurons to the cortex, so that the cortical circuit diagram differs between primates and other mammals (Fig. 1). Upper cortical layers, generated late in neurogenesis, are over-represented in the primate cerebral cortex, especially in humans⁵. Additionally, special cell types, such as spindle cells (specialized, deep-layer neurons⁶), are unique to primates. The upper-layer neurons that are so unusually common in great apes represent either locally projecting neurons—the “neurons with short axon” of Cajal—or neurons that connect the cortex to itself, but do not project out of the cortex (Fig. 1).

The cerebral cortex shows remarkable local specialization, reflected as functionally distinct cortical ‘areas’ that are essentially a map of the behaviours and capabilities most essential to each species. For example, whereas rodents show relatively larger areas that respond to odours and sensation from the whiskers, they have small areas subserving their limited vision. In contrast, primates are highly visual, with more than a dozen distinct functional areas analysing various features of a visual scene. Recent work has compared functionally homologous visual regions between humans and macaques, suggesting that some areas are quite similar, whereas other visual areas have been either added or greatly modified during the course of evolution⁷. Primates also have particularly large areas of the frontal lobes anterior to the motor cortex (prefrontal cortex), whereas prefrontal cortex is tiny in non-primates. Prefrontal areas regulate many social behaviours and are preferentially enlarged in great apes. Although it has long been thought that prefrontal cortex is especially enlarged in humans, recent work suggests that other great apes may have equivalent proportions of prefrontal cortex⁸.

The human cerebral cortex also shows functional asymmetries, with most of us being right handed and having language function preferentially localized in the left hemisphere. Chimpanzees do not show such strong asymmetry in handedness⁹, although their brains show some asymmetries in frontal and temporal lobes (which correspond to language areas in humans)¹⁰. Recent evidence suggests that the left–right asymmetries of the human cerebral cortex are accompanied by asymmetric gene expression during early fetal development¹¹, although it is not known whether asymmetries of gene expression are seen in non-human primates. There is some evidence from fossil skulls for cortical asymmetry in human predecessors as well¹².

Evolutionary mechanisms

What sorts of genetic changes underlie diverse brain shape and size? Approaches to this question have come increasingly into focus, although the answers themselves await further work. Three major mechanisms of evolutionary changes include: (1) addition or subtraction of entire genes to or from the genome; (2) alterations in levels or patterns of gene expression; and (3) alterations in the coding sequence of genes. Recent evidence suggests roles for all of these mechanisms.

The recent completion of sequencing the chimpanzee genome emphasizes the highly similar composition of the human and chimpanzee genomes¹³. There is evidence for inactivation of genes, especially many olfactory receptor genes, by their conversion into

¹Division of Neurogenetics and Howard Hughes Medical Institute, Beth Israel Deaconess Medical Center, and Department of Neurology, Harvard Medical School, Room 266, New Research Building, 77 Avenue Louis Pasteur, Boston, Massachusetts 02115, USA.

pseudogenes¹⁴. However, there is currently little evidence to suggest that the addition of novel genes is a major mechanism in human brain evolution¹³.

Recent studies suggest that human brain evolution is associated with changes in gene expression specifically within the brain as opposed to other tissues such as liver. A few studies suggest more-accelerated gene expression changes in the brain along the human lineage compared with the chimpanzee lineage¹⁵. Although the studies differ in design and principal conclusions, they share support for an increase in expression level in a subset of brain-expressed genes in the lineage leading to humans^{16,17}.

There is also accumulating evidence that some neural genes underwent important changes in their coding sequence over the course of recent brain evolution, although the proportion of neural genes that were targets of positive selection is still in debate. Genes strongly influenced by natural selection can be identified by comparing DNA changes that occur in different, closely related species, for example in different primate species. Synonymous DNA substitutions do not alter the amino acid sequence because they occur at degenerate sites in the codon (such as a CGT to CGG change, as both codons encode arginine). Because synonymous changes do not alter the biochemical properties of the encoded protein, they are usually

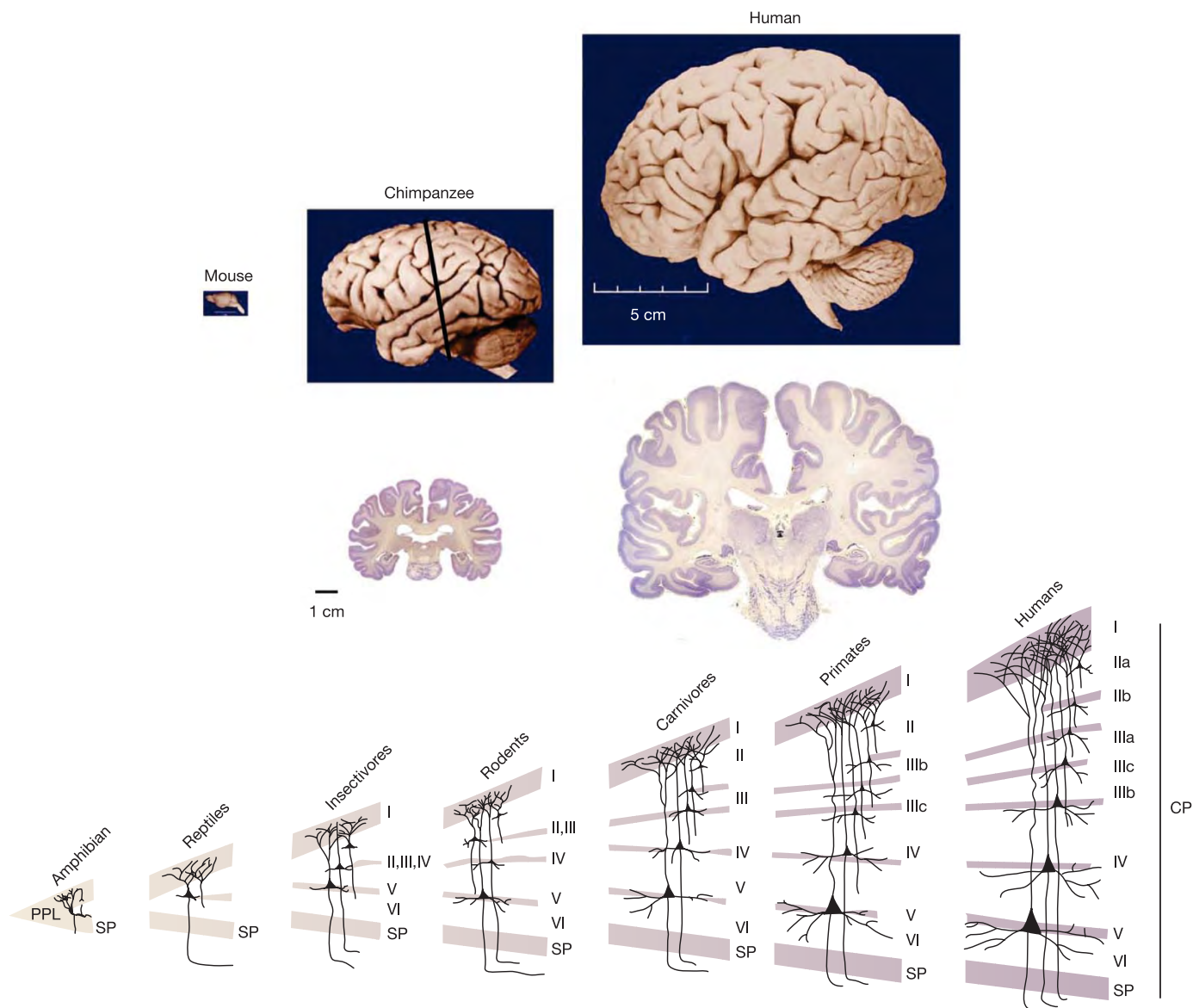


Figure 1 | Differences in cerebral cortical size are associated with differences in the cerebral cortex circuit diagram. The top panel shows side views of the brain of a rodent (mouse), a chimpanzee and a human to show relative sizes. The middle panel shows a cross-section of a human and chimpanzee brain, with the cellular composition of the cortex illustrated in the bottom panel (adapted from ref. 5). The cerebral cortex derives from two developmental cell populations: the primordial plexiform layer (PPL) and the cortical plate (CP). The primordial plexiform layer seems to be homologous to simple cortical structures in Amphibia and Reptilia, and appears first temporally during mammalian brain development. The cortical plate develops as a second population that splits the primordial plexiform

layer into two layers (layer I at the top and the subplate (SP) at the bottom; numbering follows the scheme of ref. 31). Cortical-plate-derived cortical layers are added developmentally from deeper first (VI, V) to more superficial (III, II) last. Cortical-plate-derived cortical layers are progressively elaborated in mammals with larger brains (for example, insectivores have a single layer II/III/IV that is progressively subdivided into II, III, IV, then IIa, IIb, and so on), so that humans have a larger proportion of these late-derived neurons, which project locally or elsewhere within the cortex. Images from the top and middle panels are from the Comparative Brain Atlas (<http://www.brainmuseum.org>).

evolutionarily neutral. In contrast, non-synonymous DNA changes alter the amino acid sequence. The vast majority of non-synonymous DNA changes represent disabling mutations that cause disease, hence decreasing the fitness of the organism, and so most non-synonymous DNA changes are subject to negative, or purifying, selection. In contrast, on rare occasions non-synonymous DNA changes might make the protein work slightly better, hence increasing the fitness of the organism and becoming subject to positive selection (that is, advantageous changes propagated to future generations). A ratio of non-synonymous (K_A) to synonymous (K_S) changes $\ll 1$ is typical of most proteins where change is detrimental¹⁸; rare proteins show $K_A/K_S > 1$, which can indicate positive selection.

In order to test whether genes expressed in the brain were frequent targets of positive selection in primates, one study¹⁹ analysed 200 brain-expressed genes, comparing them to 200 widely expressed genes. They compared K_A/K_S ratios between rats and mice and between humans and macaque monkeys. They concluded that genes involved in brain development or function had a higher tendency to be under positive selection between macaques and humans than between mice and rats. In contrast, systematic surveys of K_A/K_S ratios across much larger numbers of genes between chimpanzees and humans failed to show that neural genes, as a group, have higher K_A/K_S ratios than genes expressed outside of the brain between these two species^{20,21}. Analysis of the top 50 genes with the highest K_A/K_S ratios showed surprisingly few with known essential roles in the brain²⁰. Analysis of the chimpanzee genome confirms that neural genes, as a group, have much lower average K_A/K_S ratios than genes expressed outside of the brain¹³. However, the more recent study suggested that a substantial fraction of the genes with the highest K_A/K_S ratios had roles in brain development or function¹³. These studies are most easily reconciled by suggesting that a small subset of neural genes may be targets for positive selection (see below), whereas neural genes as a whole are subject to intense negative selection due to the severe disadvantages conferred by mutations that disrupt brain function.

Correlation of genetic evolution with human brain function

Whereas genome-wide analyses systematically highlight targets of positive genetic selection in the human lineage, there has been great interest in a subset of human genes that show positive evolutionary selection, and for which correlations between evolutionary patterns and gene function in humans are possible. For example, mutant alleles of *FOXP2* cause a severe disorder of articulation and speech in humans, yet subtle differences in *FOXP2* sequence between humans and non-humans show evidence of positive evolutionary selection by K_A/K_S ratio. Its involvement in speech production suggests that changes in *FOXP2* may have been important in the evolution of language^{22,23}. Furthermore, analysis of *FOXP2*'s DNA sequence in diverse human populations suggests that the gene shows unusually low sequence diversity—that is, many human populations share a common ancestral sequence at the *FOXP2* locus. This evidence for a 'selective sweep' (explained in detail in several recent reviews^{2,24}) within humans suggests that evolutionary selection on this gene may have occurred very recently in human evolution; that is, after the appearance of *Homo sapiens*.

Two genes that cause microcephaly (small cerebral cortex) also show strong evidence for positive evolutionary selection. Microcephaly reduces the human brain to 50% or less of its normal mass; that is, to about the size of the brain of chimpanzees or our pre-human ancestors. Whereas marked mutations in abnormal spindle microcephaly (encoded by the *ASPM* locus) and microcephalin (encoded by the *MCPH1* locus) cause microcephaly, both genes show strong evidence that subtler sequence changes were subject to positive selection in the lineage leading to humans (manifested by a high K_A/K_S ratio)^{25–29}. Although the precise functions of the two genes are

unknown, both are highly expressed in dividing neural precursor cells in the cerebral cortex, and available evidence suggests roles in cell proliferation. Notably, just as neurons in the upper layers of the cerebral cortex (Fig. 1) are added last during development, and are most highly elaborated in humans and great apes, these upper-layer neurons are preferentially lost in many cases of microcephaly, supporting a requirement for microcephaly genes in the formation of the upper cortical layers.

AH11, which is essential for axon pathfinding from the cortex to the spinal cord (and hence for normal coordination and gait), is another gene that causes a neurological disease when mutated, but for which subtler changes between primate species suggest positive evolutionary selection in the lineage leading to humans³⁰. Patients with *AH11* mutations not only show mental retardation, but can also show symptoms characteristic of autism, such as antisocial behaviour. This raises the intriguing possibility that evolutionary differences in *AH11* may relate not only to human patterns of gait, but potentially species-specific social behaviour.

The linkage of studies of gene function in humans with evolutionary analysis is just beginning, and is limited mainly by the rate at which the essential functional roles of genes in the human brain are elucidated. As a population, humans show many mutant alleles for every gene that has been extensively studied, so that the human population is likely to represent, to a first approximation, saturation mutagenesis, such that for each gene in the genome there is a human carrying a mutated allele for that gene. Many neurological diseases affect the very processes that define us evolutionarily as human: intelligence (mental retardation), social organization (autism and attention deficit disorder) and higher-order language (dyslexia). As the genes for these uniquely human disorders are characterized, they may give us new insight into our recent evolutionary history.

1. Ramon y Cajal, S. *Recuerdos de mi Vida* Vol. 2 *Historia de mi Labour Científica* 345–346 (Moya, Madrid, 1917).
2. Carroll, S. B. Genetics and the making of *Homo sapiens*. *Nature* **422**, 849–857 (2003).
3. Kornack, D. R. & Rakic, P. Changes in cell-cycle kinetics during the development and evolution of primate neocortex. *Proc. Natl Acad. Sci. USA* **95**, 1242–1246 (1998).
4. Takahashi, T., Nowakowski, R. S. & Caviness, V. S. Jr. The cell cycle of the pseudostratified ventricular epithelium of the embryonic murine cerebral wall. *J. Neurosci.* **15**, 6046–6057 (1995).
5. Marin-Padilla, M. Ontogenesis of the pyramidal cell of the mammalian neocortex and developmental cytoarchitectonics: a unifying theory. *J. Comp. Neurol.* **321**, 223–240 (1992).
6. Allman, J., Hakeem, A. & Watson, K. Two phylogenetic specializations in the human brain. *Neuroscientist* **8**, 335–346 (2002).
7. Orban, G. A., Van Essen, D. & Vanduffel, W. Comparative mapping of higher visual areas in monkeys and humans. *Trends Cogn. Sci.* **8**, 315–324 (2004).
8. Semendefer, K., Lu, A., Schenker, N. & Damasio, H. Humans and great apes share a large frontal cortex. *Nature Neurosci.* **5**, 272–276 (2002).
9. Hopkins, W. D. & Cantalupo, C. Handedness in chimpanzees (*Pan troglodytes*) is associated with asymmetries of the primary motor cortex but not with homologous language areas. *Behav. Neurosci.* **118**, 1176–1183 (2004).
10. Cantalupo, C. & Hopkins, W. D. Asymmetric Broca's area in great apes. *Nature* **414**, 505 (2001).
11. Sun, T. et al. Early asymmetry of gene transcription in embryonic human left and right cerebral cortex. *Science* **308**, 1794–1798 (2005).
12. Broadfield, D. C. et al. Endocast of *Sambungmacan 3* (Sm 3): a new *Homo erectus* from Indonesia. *Anat. Rec.* **262**, 369–379 (2001).
13. The Chimpanzee Sequencing and Analysis Consortium. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* doi:10.1038/nature04072 (this issue).
14. Gilad, Y., Man, O. & Glusman, G. A comparison of the human and chimpanzee olfactory receptor gene repertoires. *Genome Res.* **15**, 224–230 (2005).
15. Enard, W. et al. Intra- and interspecific variation in primate gene expression patterns. *Science* **296**, 340–343 (2002).
16. Caceres, M. et al. Elevated gene expression levels distinguish human from non-human primate brains. *Proc. Natl Acad. Sci. USA* **100**, 13030–13035 (2003).
17. Uddin, M. et al. Sister grouping of chimpanzees and humans as revealed by genome-wide phylogenetic analysis of brain gene expression profiles. *Proc. Natl Acad. Sci. USA* **101**, 2957–2962 (2004).

18. Goldman, N. & Yang, Z. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol. Biol. Evol.* **11**, 725–736 (1994).
19. Dorus, S. *et al.* Accelerated evolution of nervous system genes in the origin of *Homo sapiens*. *Cell* **119**, 1027–1040 (2004).
20. Nielsen, R. *et al.* A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biol.* **3**, e170 (2005).
21. Clark, A. G. *et al.* Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios. *Science* **302**, 1960–1963 (2003).
22. Lai, C. S., Fisher, S. E., Hurst, J. A., Vargha-Khadem, F. & Monaco, A. P. A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature* **413**, 519–523 (2001).
23. Enard, W. *et al.* Molecular evolution of *FOXP2*, a gene involved in speech and language. *Nature* **418**, 869–872 (2002).
24. Gilbert, S. L., Dobyns, W. B. & Lahn, B. T. Genetic links between brain development and brain evolution. *Nature Rev. Genet.* **6**, 581–590 (2005).
25. Bond, J. *et al.* *ASPM* is a major determinant of cerebral cortical size. *Nature Genet.* **32**, 316–320 (2002).
26. Evans, P. D., Anderson, J. R., Vallender, E. J., Choi, S. S. & Lahn, B. T. Reconstructing the evolutionary history of microcephalin, a gene controlling human brain size. *Hum. Mol. Genet.* **13**, 1139–1145 (2004).
27. Evans, P. D. *et al.* Adaptive evolution of *ASPM*, a major determinant of cerebral cortical size in humans. *Hum. Mol. Genet.* **13**, 489–494 (2004).
28. Kouprina, N. *et al.* Accelerated evolution of the *ASPM* gene controlling brain size begins prior to human brain expansion. *PLoS Biol.* **2**, E126 (2004).
29. Zhang, J. Evolution of the human *ASPM* gene, a major determinant of brain size. *Genetics* **165**, 2063–2070 (2003).
30. Ferland, R. J. *et al.* Abnormal cerebellar development and axonal decussation due to mutations in *AH11* in Joubert syndrome. *Nature Genet.* **36**, 1008–1013 (2004).
31. Marin-Padilla, M. Dual origin of the mammalian neocortex and evolution of the cortical plate. *Anat. Embryol.* **152**, 109–126 (1978).

Acknowledgements This work was supported by grants from the NINDS and Cure Autism Now. We thank M. Ruvolo and D. Reich for comments on an earlier version of this manuscript, and J. DeFilipe for the translation of the Cajal quotation. Owing to space limitations we were unable to cite directly some of the relevant work in this field. C.A.W. is an Investigator of the Howard Hughes Medical Institute.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to C.A.W. (cwalsh@bidmc.harvard.edu).

Initial sequence of the chimpanzee genome and comparison with the human genome

The Chimpanzee Sequencing and Analysis Consortium*

Here we present a draft genome sequence of the common chimpanzee (*Pan troglodytes*). Through comparison with the human genome, we have generated a largely complete catalogue of the genetic differences that have accumulated since the human and chimpanzee species diverged from our common ancestor, constituting approximately thirty-five million single-nucleotide changes, five million insertion/deletion events, and various chromosomal rearrangements. We use this catalogue to explore the magnitude and regional variation of mutational forces shaping these two genomes, and the strength of positive and negative selection acting on their genes. In particular, we find that the patterns of evolution in human and chimpanzee protein-coding genes are highly correlated and dominated by the fixation of neutral and slightly deleterious alleles. We also use the chimpanzee genome as an outgroup to investigate human population genetics and identify signatures of selective sweeps in recent human evolution.

More than a century ago Darwin¹ and Huxley² posited that humans share recent common ancestors with the African great apes. Modern molecular studies have spectacularly confirmed this prediction and have refined the relationships, showing that the common chimpanzee (*Pan troglodytes*) and bonobo (*Pan paniscus* or pygmy chimpanzee) are our closest living evolutionary relatives³. Chimpanzees are thus especially suited to teach us about ourselves, both in terms of their similarities and differences with human. For example, Goodall's pioneering studies on the common chimpanzee revealed startling behavioural similarities such as tool use and group aggression^{4,5}. By contrast, other features are obviously specific to humans, including habitual bipedality, a greatly enlarged brain and complex language⁵. Important similarities and differences have also been noted for the incidence and severity of several major human diseases⁶.

Genome comparisons of human and chimpanzee can help to reveal the molecular basis for these traits as well as the evolutionary forces that have moulded our species, including underlying mutational processes and selective constraints. Early studies sought to draw inferences from sets of a few dozen genes^{7–9}, whereas recent studies have examined larger data sets such as protein-coding exons¹⁰, random genomic sequences^{11,12} and an entire chimpanzee chromosome¹³.

Here we report a draft sequence of the genome of the common chimpanzee, and undertake comparative analyses with the human genome. This comparison differs fundamentally from recent comparative genomic studies of mouse, rat, chicken and fish^{14–17}. Because these species have diverged substantially from the human lineage, the focus in such studies is on accurate alignment of the genomes and recognition of regions of unusually high evolutionary conservation to pinpoint functional elements. Because the chimpanzee lies at such a short evolutionary distance with respect to human, nearly all of the bases are identical by descent and sequences can be readily aligned except in recently derived, large repetitive regions. The focus thus turns to differences rather than similarities. An observed difference at a site nearly always represents a single event, not multiple indepen-

dent changes over time. Most of the differences reflect random genetic drift, and thus they hold extensive information about mutational processes and negative selection that can be readily mined with current analytical techniques. Hidden among the differences is a minority of functionally important changes that underlie the phenotypic differences between the two species. Our ability to distinguish such sites is currently quite limited, but the catalogue of human–chimpanzee differences opens this issue to systematic investigation for the first time. We would also hope that, in elaborating the few differences that separate the two species, we will increase pressure to save chimpanzees and other great apes in the wild.

Our results confirm many earlier observations, but notably challenge some previous claims based on more limited data. The genome-wide data also allow some questions to be addressed for the first time. (Here and throughout, we refer to chimpanzee–human comparison as representing hominids and mouse–rat comparison as representing murids—of course, each pair covers only a subset of the clade.) The main findings include:

- Single-nucleotide substitutions occur at a mean rate of 1.23% between copies of the human and chimpanzee genome, with 1.06% or less corresponding to fixed divergence between the species.
- Regional variation in nucleotide substitution rates is conserved between the hominid and murid genomes, but rates in subtelomeric regions are disproportionately elevated in the hominids.
- Substitutions at CpG dinucleotides, which constitute one-quarter of all observed substitutions, occur at more similar rates in male and female germ lines than non-CpG substitutions.
- Insertion and deletion (indel) events are fewer in number than single-nucleotide substitutions, but result in ~1.5% of the euchromatic sequence in each species being lineage-specific.
- There are notable differences in the rate of transposable element insertions: short interspersed elements (SINEs) have been threefold more active in humans, whereas chimpanzees have acquired two new families of retroviral elements.

*Lists of participants and affiliations appear at the end of the paper.

- Orthologous proteins in human and chimpanzee are extremely similar, with ~29% being identical and the typical orthologue differing by only two amino acids, one per lineage.
- The normalized rates of amino-acid-altering substitutions in the hominid lineages are elevated relative to the murid lineages, but close to that seen for common human polymorphisms, implying that positive selection during hominid evolution accounts for a smaller fraction of protein divergence than suggested in some previous reports.
- The substitution rate at silent sites in exons is lower than the rate at nearby intronic sites, consistent with weak purifying selection on silent sites in mammals.
- Analysis of the pattern of human diversity relative to hominid divergence identifies several loci as potential candidates for strong selective sweeps in recent human history.

In this paper, we begin with information about the generation, assembly and evaluation of the draft genome sequence. We then explore overall genome evolution, with the aim of understanding mutational processes at work in the human genome. We next focus on the evolution of protein-coding genes, with the aim of characterizing the nature of selection. Finally, we briefly discuss initial insights into human population genetics.

In recognition of its strong community support, we will refer to chimpanzee chromosomes using the orthologous numbering nomenclature proposed by ref. 18, which rennumbers the chromosomes of the great apes from the International System for Human Cytogenetic Nomenclature (ISCN; 1978) standard to directly correspond to their human orthologues, using the terms 2A and 2B for the two ape chromosomes corresponding to human chromosome 2.

Genome sequencing and assembly

We sequenced the genome of a single male chimpanzee (Clint; Yerkes pedigree number C0471; Supplementary Table S1), a captive-born descendant of chimpanzees from the West Africa subspecies *Pan troglodytes verus*, using a whole-genome shotgun (WGS) approach^{19,20}. The data were assembled using both the PCAP and ARACHNE programs^{21,22} (see Supplementary Information 'Genome sequencing and assembly' and Supplementary Tables S2–S6). The former was a *de novo* assembly, whereas the latter made limited use of human genome sequence (NCBI build 34)^{23,24} to facilitate and confirm contig linking. The ARACHNE assembly has slightly greater continuity (Table 1) and was used for analysis in this paper. The draft genome assembly—generated from ~3.6-fold sequence redundancy of the autosomes and ~1.8-fold redundancy of both sex chromosomes—covers ~94% of the chimpanzee genome with >98% of the sequence in high-quality bases. A total of 50% of the sequence (N50) is contained in contigs of length greater than 15.7 kilobases (kb) and supercontigs of length greater than 8.6 megabases (Mb). The assembly represents a consensus of two haplotypes, with one allele from each heterozygous position arbitrarily represented in the sequence.

Assessment of quality and coverage. The chimpanzee genome assembly was subjected to rigorous quality assessment, based on comparison to finished chimpanzee bacterial artificial chromosomes (BACs) and to the human genome (see Supplementary Information

'Genome sequencing and assembly' and Supplementary Tables S7–S16).

Nucleotide-level accuracy is high by several measures. About 98% of the chimpanzee genome sequence has quality scores²⁵ of at least 40 (Q40), corresponding to an error rate of $\leq 10^{-4}$. Comparison of the WGS sequence to 1.3 Mb of finished BACs from the sequenced individual is consistent with this estimate, giving a high-quality discrepancy rate of 3×10^{-4} substitutions and 2×10^{-4} indels, which is no more than expected given the heterozygosity rate (see below), as 50% of the polymorphic alleles in the WGS sequence will differ from the single-haplotype BACs. Comparison of protein-coding regions aligned between the WGS sequence, the recently published sequence of chimpanzee chromosome 21 (ref. 13; formerly chromosome 22 (ref. 18)) and the human genome also revealed no excess of substitutions in the WGS sequence (see Supplementary Information 'Genome sequencing and assembly'). Thus, by restricting our analysis to high-quality bases, the nucleotide-level accuracy of the WGS assembly is essentially equal to that of 'finished' sequence.

Structural accuracy is also high based on comparison with finished BACs from the primary donor and other chimpanzees, although the relatively low level of sequence redundancy limits local contiguity. On the basis of comparisons with the primary donor, some small supercontigs (most <5 kb) have not been positioned within large supercontigs (~1 event per 100 kb); these are not strictly errors but nonetheless affect the utility of the assembly. There are also small, undetected overlaps (all <1 kb) between consecutive contigs (~1.2 events per 100 kb) and occasional local misordering of small contigs (~0.2 events per 100 kb). No misoriented contigs were found. Comparison with the finished chromosome 21 sequence yielded similar discrepancy rates (see Supplementary Information 'Genome sequencing and assembly').

The most problematic regions are those containing recent segmental duplications. Analysis of BAC clones from duplicated ($n = 75$) and unique ($n = 28$) regions showed that the former tend to be fragmented into more contigs (1.6-fold) and more supercontigs (3.2-fold). Discrepancies in contig order are also more frequent in duplicated than unique regions (~0.4 versus ~0.1 events per 100 kb). The rate is twofold higher in duplicated regions with the highest sequence identity (>98%). If we restrict the analysis to older duplications ($\leq 98\%$ identity) we find fewer assembly problems: 72% of those that can be mapped to the human genome are shared as duplications in both species. These results are consistent with the described limitations of current WGS assembly for regions of segmental duplication²⁶. Detailed analysis of these rapidly changing regions of the genome is being performed with more directed approaches²⁷.

Chimpanzee polymorphisms. The draft sequence of the chimpanzee genome also facilitates genome-wide studies of genetic diversity among chimpanzees, extending recent work^{28–31}. We sequenced and analysed sequence reads from the primary donor, four other West African and three central African chimpanzees (*Pan troglodytes troglodytes*) to discover polymorphic positions within and between these individuals (Supplementary Table S17).

A total of 1.66 million high-quality single-nucleotide polymorphisms (SNPs) were identified, of which 1.01 million are heterozygous within the primary donor, Clint. Heterozygosity rates were estimated to be 9.5×10^{-4} for Clint, 8.0×10^{-4} among West African chimpanzees and 17.6×10^{-4} among central African chimpanzees, with the variation between West and central African chimpanzees being 19.0×10^{-4} . The diversity in West African chimpanzees is similar to that seen for human populations³², whereas the level for central African chimpanzees is roughly twice as high.

The observed heterozygosity in Clint is broadly consistent with West African origin, although there are a small number of regions of distinctly higher heterozygosity. These may reflect a small amount of central African ancestry, but more likely reflect undetected regions of segmental duplications present only in chimpanzees.

Table 1 | Chimpanzee assembly statistics

Assembler	PCAP	ARACHNE
Major contigs*	400,289	361,782
Contig length (kb; N50)†	13.3	15.7
Supercontigs	67,734	37,846
Supercontig length (Mb; N50)	2.3	8.6
Sequence redundancy: all bases (Q20)	$5.0 \times (3.6 \times)$	$4.3 \times (3.6 \times)$
Physical redundancy	20.7	19.8
Consensus bases (Gb)	2.7	2.7

*Contigs >1 kb.

†N50 length is the size x such that 50% of the assembly is in units of length at least x .

Genome evolution

We set out to study the mutational events that have shaped the human and chimpanzee genomes since their last common ancestor. We explored changes at the level of single nucleotides, small insertions and deletions, interspersed repeats and chromosomal rearrangements. The analysis is nearly definitive for the smallest changes, but is more limited for larger changes, particularly lineage-specific segmental duplications, owing to the draft nature of the genome sequence.

Nucleotide divergence. Best reciprocal nucleotide-level alignments of the chimpanzee and human genomes cover ~2.4 gigabases (Gb) of high-quality sequence, including 89 Mb from chromosome X and 7.5 Mb from chromosome Y.

Genome-wide rates. We calculate the genome-wide nucleotide divergence between human and chimpanzee to be 1.23%, confirming recent results from more limited studies^{12,33,34}. The differences between one copy of the human genome and one copy of the chimpanzee genome include both the sites of fixed divergence between the species and some polymorphic sites within each species. By correcting for the estimated coalescence times in the human and chimpanzee populations (see Supplementary Information ‘Genome evolution’), we estimate that polymorphism accounts for 14–22% of the observed divergence rate and thus that the fixed divergence is ~1.06% or less.

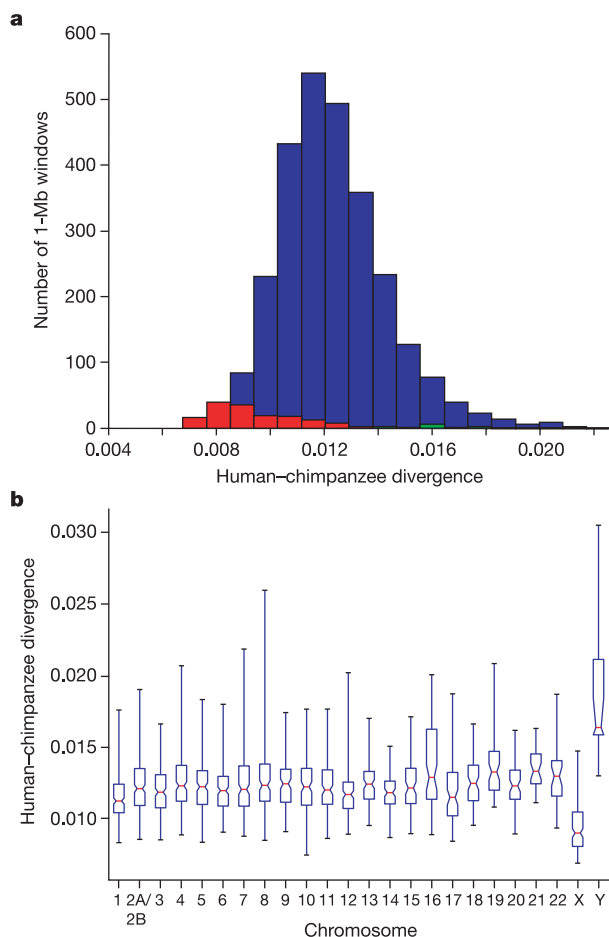


Figure 1 | Human-chimpanzee divergence in 1-Mb segments across the genome. **a**, Distribution of divergence of the autosomes (blue), the X chromosome (red) and the Y chromosome (green). **b**, Distribution of variation by chromosome, shown as a box plot. The edges of the box correspond to quartiles; the notches to the standard error of the median; and the vertical bars to the range. The X and Y chromosomes are clear outliers, but there is also high local variation within each of the autosomes.

Nucleotide divergence rates are not constant across the genome, as has been seen in comparisons of the human and murid genomes^{16,17,24,35,36}. The average divergence in 1-Mb segments fluctuates with a standard deviation of 0.25% (coefficient of variation = 0.20), which is much greater than the 0.02% expected assuming a uniform divergence rate (Fig. 1a; see also Supplementary Fig. S1).

Regional variation in divergence could reflect local variation in either mutation rate or other evolutionary forces. Among the latter, one important force is genetic drift, which can cause substantial differences in divergence time across loci when comparing closely related species, as the divergence time for orthologues is the sum of two terms: t_1 , the time since speciation, and t_2 , the coalescence time for orthologues within the common ancestral population³⁷. Whereas t_1 is constant across loci (~6–7 million years³⁸), t_2 is a random variable that fluctuates across loci (with a mean that depends on population size and here may be on the order of 1–2 million years³⁹). However, because of historical recombination, the characteristic scale of such fluctuations will be on the order of tens of kilobases, which is too small to account for the variation observed for 1-Mb regions⁴⁰ (see Supplementary Information ‘Genome evolution’). Other potential evolutionary forces are positive or negative selection. Although it is more difficult to quantify the expected contributions of selection in the ancestral population^{41–43}, it is clear that the effects would have to be very strong to explain the large-scale variation observed across mammalian genomes^{16,44}. There is tentative evidence from in-depth analysis of divergence and diversity that natural selection is not the major contributor to the large-scale patterns of genetic variability in humans^{45–47}. For these reasons, we suggest that the large-scale variation in the human–chimpanzee divergence rate primarily reflects regional variation in mutation rate.

Chromosomal variation in divergence rate. Variation in divergence rate is evident even at the level of whole chromosomes (Fig. 1b). The most striking outliers are the sex chromosomes, with a mean divergence of 1.9% for chromosome Y and 0.94% for chromosome X. The likely explanation is a higher mutation rate in the male compared with female germ line⁴⁸. Indeed, the ratio of the male/female mutation rates (denoted α) can be estimated by comparing the divergence rates among the sex chromosomes and the autosomes and correcting for ancestral polymorphism as a function of population size of the most recent common ancestor (MRCA; see Supplementary Information ‘Genome evolution’). Estimates for α range from 3 to 6, depending on the chromosomes compared and the assumed ancestral population size (Supplementary Table S18). This is significantly higher than recent estimates of α for the murids (~1.9) (ref. 17) and resolves a recent controversy based on smaller data sets^{12,24,49,50}.

The higher mutation rate in the male germ line is generally attributed to the 5–6-fold higher number of cell divisions undergone by male germ cells⁴⁸. We reasoned that this would affect mutations resulting from DNA replication errors (the rate should scale with the number of cell divisions) but not mutations resulting from DNA damage such as deamination of methyl CpG to TpG (the rate should scale with time). Accordingly, we calculated α separately for CpG sites, obtaining a value of ~2 from the comparison of rates between autosomes and chromosome X. This intermediate value is a composite of the rates of CpG loss and gain, and is consistent with roughly equal rates of CpG to TpG transitions in the male and female germ line^{51,52}.

Significant variation in divergence rates is also seen among autosomes (Fig. 1b; $P < 3 \times 10^{-15}$, Kruskal–Wallis test over 1-Mb windows), confirming earlier observations based on low-coverage WGS sampling¹². Additional factors thus influence the rate of divergence between chimpanzee and human chromosomes. These factors are likely to act at length scales significantly shorter than a chromosome, because the standard deviation across autosomes (0.21%) is comparable to the standard deviation seen in 1-Mb windows across the genome (0.13–0.35%). We therefore sought to

understand local factors that contribute to variation in divergence rate.

Contribution of CpG dinucleotides. Sites containing CpG dinucleotides in either species show a substantially elevated divergence rate of 15.2% per base; they account for 25.2% of all substitutions while constituting only 2.1% of all aligned bases. The divergence at CpG sites represents both the loss of ancestral CpGs and the creation of new CpGs. The former process is known to occur at a rapid rate per base due to frequent methylation of cytosines in a CpG context and their frequent deamination^{53,54}, whereas the latter process probably proceeds at a rate more typical of other nucleotide substitutions. Assuming that loss and creation of CpG sites are close to equilibrium, the mutation rate for bases in a CpG dinucleotide must be 10–12-fold higher than for other bases (see Supplementary Information ‘Genome evolution’ and ref. 51).

Because of the high rate of CpG substitutions, regional divergence rates would be expected to correlate with regional CpG density. CpG density indeed varies across 1-Mb windows (mean = 2.1%, coefficient of variation = 0.44 compared with 0.0093 expected under a Poisson distribution), but only explains 4% of the divergence rate variance. In fact, regional CpG and non-CpG divergence is highly correlated ($r = 0.88$; Supplementary Fig. S2), suggesting that higher-order effects modulate the rates of two very different mutation processes (see also ref. 47).

Increased divergence in distal regions. The most striking regional pattern is a consistent increase in divergence towards the ends of most chromosomes (Fig. 2). The terminal 10 Mb of chromosomes (including distal regions and proximal regions of acrocentric chromosomes) averages 15% higher divergence than the rest of the genome (Mann–Whitney U -test; $P < 10^{-30}$), with a sharp increase towards the telomeres. The phenomenon correlates better with physical distance than relative position along the chromosomes and may partially explain why smaller chromosomes tend to have higher divergence (Supplementary Fig. S3; see also ref. 15). These observations suggest that large-scale chromosomal structure, directly or indirectly, influences regional divergence patterns. The cause of this effect is unclear, but these regions (~15% of the genome) are

notable in having high local recombination rate, high gene density and high G + C content.

Correlation with chromosome banding. Another interesting pattern is that divergence increases with the intensity of Giemsa staining in cytogenetically defined chromosome bands, with the regions corresponding to Giemsa dark bands (G bands) showing 10% higher divergence than the genome-wide average (Mann–Whitney U -test; $P < 10^{-14}$) (see Fig. 2). In contrast to terminal regions, these regions (17% of the genome) tend to be gene poor, (G + C)-poor and low in recombination^{55,56}. The elevated divergence seen in two such different types of regions suggests that multiple mechanisms are at work, and that no single known factor, such as G + C content or recombination rate, is an adequate predictor of regional variation in the mammalian genome by itself (Fig. 3). Elucidation of the relative contributions of these and other mechanisms will be important for formulating accurate models for population genetics, natural selection, divergence times and the evolution of genome-wide sequence composition⁵⁷.

Correlation with regional variation in the murid genome. Given that sequence divergence shows regional variation in both hominids (human–chimpanzee) and murids (mouse–rat), we asked whether the regional rates are positively correlated between orthologous regions. Such a correlation would suggest that the divergence rate is driven, in part, by factors that have been conserved over the ~75 million years since rodents, humans and apes shared a common ancestor. Comparative analysis of the human and murid genomes has suggested such a correlation^{58–60}, but the chimpanzee sequence provides a direct opportunity to compare independent evolutionary processes between two mammalian clades.

We compared the local divergence rates in hominids and murids across major orthologous segments in the respective genomes (Fig. 4). For orthologous segments that are non-distal in both hominids and murids, there is a strong correlation between the divergence rates ($r = 0.5$, $P < 10^{-11}$). In contrast, orthologous segments that are centred within 10 Mb of a hominid telomere have disproportionately high divergence rates and G + C content relative to the murids (Mann–Whitney U -test; $P < 10^{-11}$ and

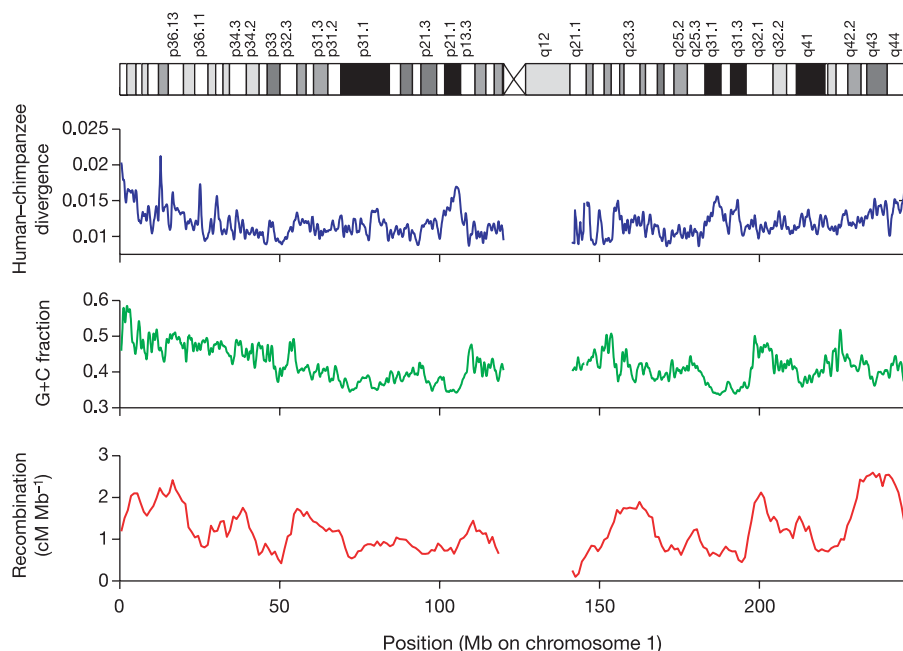


Figure 2 | Regional variation in divergence rates. Human–chimpanzee divergence (blue), G + C content (green) and human recombination rates¹⁷³ (red) in sliding 1-Mb windows for human and chimpanzee chromosome 1. Divergence and G + C content are noticeably elevated near the 1p telomere,

a trend that holds for most subtelomeric regions (see text). Internally on the chromosome, regions of low G + C content and high divergence often correspond to the dark G bands.

$P < 10^{-4}$), implying that the elevation in these regions is, at least partially, lineage specific. The same general effect is observed (albeit less pronounced) if CpG dinucleotides are excluded (Supplementary Fig. S4). Increased divergence and G + C content might be explained by 'biased gene conversion'⁶¹ due to the high hominid recombination rates in these distal regions. Segments that are distal in murids do not show elevated divergence rates, which is consistent with this model, because the recombination rates of distal regions are not as elevated in mouse and rat⁶².

Taken together, these observations suggest that sequence divergence rate is influenced by both conserved factors (stable across mammalian evolution) and lineage-specific factors (such as proximity to the telomere or recombination rate, which may change with chromosomal rearrangements).

Insertions and deletions. We next studied the indel events that have occurred in the human and chimpanzee lineages by aligning the genome sequences to identify length differences. We will refer below to all events as insertions relative to the other genome, although they may represent insertions or deletions relative to the genome of the common ancestor.

The observable insertions fall into two classes: (1) 'completely covered' insertions, occurring within continuous sequence in both species; and (2) 'incompletely covered' insertions, occurring within sequence containing one or more gaps in the chimpanzee, but revealed by a clear discrepancy between the species in sequence length. Different methods are needed for reliable identification of modest-sized insertions (1 base to 15 kb) and large insertions (>15 kb), with the latter only being reliably identifiable in the human genome (see Supplementary Information 'Genome evolution').

The analysis of modest-sized insertions reveals ~32 Mb of human-specific sequence and ~35 Mb of chimpanzee-specific sequence, contained in ~5 million events in each species (Supplementary Information 'Genome evolution' and Supplementary Fig. S5). Nearly all of the human insertions are completely covered, whereas only half of the chimpanzee insertions are completely covered. Analysis of the completely covered insertions shows that the vast majority are small (45% of events cover only 1 base pair (bp), 96% are <20 bp and 98.6% are <80 bp), but that the largest few contain most of the

sequence (with the ~70,000 indels larger than 80 bp comprising 73% of the affected base pairs) (Fig. 5). The latter indels >80 bp fall into three categories: (1) about one-quarter are newly inserted transposable elements; (2) more than one-third are due to microsatellite and satellite sequences; (3) and the remainder are assumed to be mostly deletions in the other genome.

The analysis of larger insertions (>15 kb) identified 163 human regions containing 8.3 Mb of human-specific sequence in total (Fig. 6). These cases include 34 regions that involve exons from known genes, which are discussed in a subsequent section. Although we have no direct measure of large insertions in the chimpanzee genome, it appears likely that the situation is similar.

On the basis of this analysis, we estimate that the human and chimpanzee genomes each contain 40–45 Mb of species-specific euchromatic sequence, and the indel differences between the genomes thus total ~90 Mb. This difference corresponds to ~3% of both genomes and dwarfs the 1.23% difference resulting from nucleotide substitutions; this confirms and extends several recent studies^{63–67}. Of course, the number of indel events is far fewer than the number of substitution events (~5 million compared with ~35 million, respectively).

Transposable element insertions. We next used the catalogue of lineage-specific transposable element copies to compare the activity of transposons in the human and chimpanzee lineages (Table 2).

Endogenous retroviruses. Endogenous retroviruses (ERVs) have become all but extinct in the human lineage, with only a single retrovirus (human endogenous retrovirus K (HERV-K)) still active²⁴. HERV-K was found to be active in both lineages, with at least 73 human-specific insertions (7 full length and 66 solo long terminal repeats (LTRs)) and at least 45 chimpanzee-specific insertions (1 full length and 44 solo LTRs). A few other ERV classes persisted in the human genome beyond the human–chimpanzee split, leaving ~9 human-specific insertions (all solo LTRs, including five HERV9 elements) before dying out.

Against this background, it was surprising to find that the chimpanzee genome has two active retroviral elements (PtERV1 and PtERV2) that are unlike any older elements in either genome;

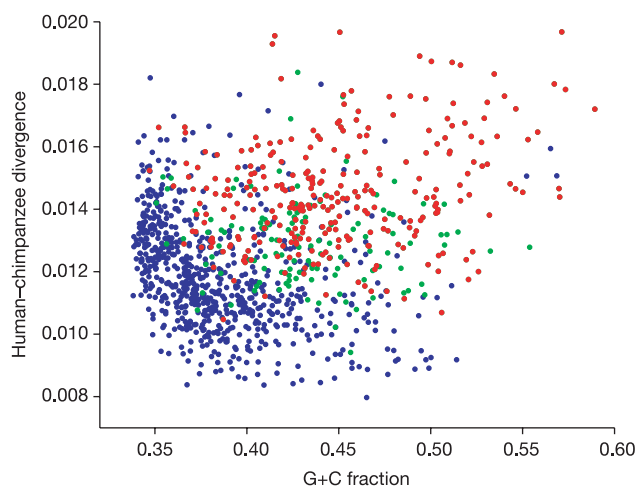


Figure 3 | Divergence rates versus G + C content for 1-Mb segments across the autosomes. Conditional on recombination rate, the relationship between divergence and G + C content varies. In regions with recombination rates less than 0.8 cM Mb^{-1} (blue), there is an inverse relationship, where high divergence regions tend to be (G + C)-poor and low divergence regions tend to be (G + C)-rich. In regions with recombination rates greater than 2.0 cM Mb^{-1} , whether within 10 Mb (red) or proximal (green) of chromosome ends, both divergence and G + C content are uniformly high.

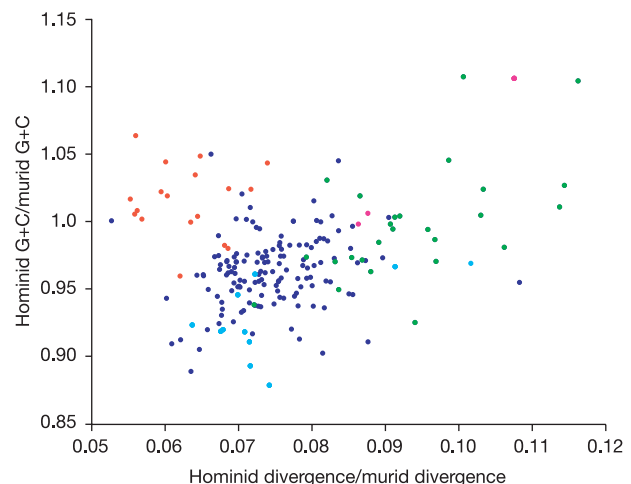


Figure 4 | Disproportionately elevated divergence and G + C content near hominid telomeres. Scatter plot of the ratio of human–chimpanzee divergence over mouse–rat divergence versus the ratio of human G + C content over mouse G + C content across 199 syntenic blocks for which more than 1 Mb of sequence could be aligned between all four species. Blocks for which the centre is within 10 Mb of a telomere in hominids only (green) or in hominids and murids (magenta), but not in murids only (light blue), show a significant trend towards higher ratios than internal blocks (dark blue). Blocks on the X chromosome (red) tend to show a lower divergence ratio than autosomal blocks, consistent with a smaller difference between autosomal and X divergence in murids than in hominids (lower α).

these must have been introduced by infection of the chimpanzee germ line. The smaller family (PtERV2) has only a few dozen copies, which nonetheless represent multiple (~ 5 – 8) invasions, because the sequence differences among reconstructed subfamilies are too great ($\sim 8\%$) to have arisen by mutation since divergence from human. It is closely related to a baboon endogenous retrovirus (BaEV, 88% ORF2 product identity) and a feline endogenous virus (ECE-1, 86% ORF2 product identity). The larger family (PtERV1) is more homogeneous and has over 200 copies. Whereas older ERVs, like HERV-K, are primarily represented by solo LTRs resulting from LTR–LTR recombination, more than half of the PtERV1 copies are still full length, probably reflecting the young age of the elements. PtERV1-like elements are present in the rhesus monkey, olive baboon and African great apes but not in human, orang-utan or gibbon, suggesting separate germline invasions in these species⁶⁸.

Higher Alu activity in humans. SINE (Alu) elements have been threefold more active in humans than chimpanzee ($\sim 7,000$ compared with $\sim 2,300$ lineage-specific copies in the aligned portion), refining the rather broad range (2–7-fold) estimated in smaller studies^{13,67,69}. Most chimpanzee-specific elements belong to a subfamily (AluYc1) that is very similar to the source gene in the common ancestor. By contrast, most human-specific Alu elements belong to two new subfamilies (AluYa5 and AluYb8) that have evolved since the chimpanzee–human divergence and differ substantially from the ancestral source gene⁶⁹. It seems likely that the resurgence of Alu elements in humans is due to these potent new source genes. However, based on an examination of available finished sequence, the baboon shows a 1.6-fold higher Alu activity relative to human new insertions, suggesting that there may also have been a general decline in activity in the chimpanzee⁶⁷.

Some of the human-specific Alu elements are highly diverged (92 with $>5\%$ divergence), which would seem to suggest that they are much older than the human–chimpanzee split. Possible explanations include: gene conversion by nearby older elements; processed pseudogenes arising from a spurious transcription of an older element; precise excision from the chimpanzee genome; or high local mutation rate. In any case, the presence of such anomalies suggests that caution is warranted in the use of single-repeat elements as homoplasy-free phylogenetic markers.

New Alu elements target (A + T)-rich DNA in human and chimpanzee genomes. Older SINE elements are preferentially found in gene-rich,

(G + C)-rich regions, whereas younger SINE elements are found in gene-poor, (A + T)-rich regions where long interspersed element (LINE)-1 (L1) copies also accumulate^{24,70}. The latter distribution is consistent with the fact that Alu retrotransposition is mediated by L1 (ref. 71). Murid genomes revealed no change in SINE distribution with age¹⁷.

The human pattern might reflect either preferential retention of SINEs in (G + C)-rich regions, due to selection or mutation bias, or a recent change in Alu insertion preferences. With the availability of the chimpanzee genome, it is possible to classify the youngest Alu copies more accurately and thus begin to distinguish these possibilities.

Analysis shows that lineage-specific SINEs in both human and chimpanzee are biased towards (A + T)-rich regions, as opposed to even the most recent copies in the MRCA (Fig. 7). This indicates that SINEs are indeed preferentially retained in (G + C)-rich DNA, but comparison with a more distant primate is required to formally rule out the possibility that the insertion bias of SINEs did not change just before speciation.

Equal activity of L1 in both species. The human and chimpanzee genomes both show $\sim 2,000$ lineage-specific L1 elements, contrary to previous estimates based on small samples that L1 activity is 2–3-fold higher in chimpanzee⁷².

Transcription from L1 source genes can sometimes continue into 3' flanking regions, which can then be co-transposed^{73,74}. Human–chimpanzee comparison revealed that $\sim 15\%$ of the species-specific insertions appear to have carried with them at least 50 bp of flanking sequence (followed by a poly(A) tail and a target site duplication). In principle, incomplete reverse transcription could result in insertions of the flanking sequence only (without any L1 sequence), mobilizing gene elements such as exons, but we found no evidence of this.

Retrotransposed gene copies. The L1 machinery also mediates retrotransposition of host messenger RNAs, resulting in many intronless (processed) pseudogenes in the human genome^{75–77}. We identified 163 lineage-specific retrotransposed gene copies in human and 246 in chimpanzee (Supplementary Table S19). Correcting for incomplete sequence coverage of the chimpanzee genome, we estimate that there are ~ 200 and ~ 300 processed gene copies in human and chimpanzee, respectively. Processed genes thus appear to have arisen at a rate of ~ 50 per million years since the divergence of human and chimpanzee; this is lower than the estimated rate for early primate evolution⁷⁵, perhaps reflecting the overall decrease in L1 activity. As expected⁷⁸, ribosomal protein genes constitute the largest class in both species. The second largest class in chimpanzee corresponds to zinc finger C2H2 genes, which are not a major class in the human genome.

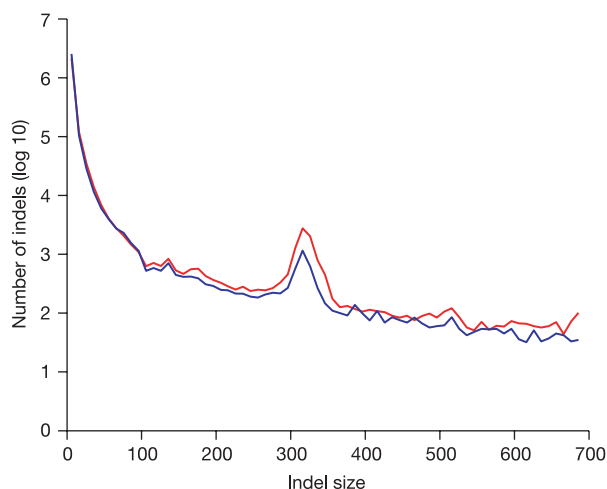


Figure 5 | Length distribution of small indel events, as determined using bounded sequence gaps. Sequences present in chimpanzee but not in human (blue) or present in human but not in chimpanzee (red) are shown. The prominent spike around 300 nucleotides corresponds to SINE insertion events. Most of the indels are smaller than 20 bp, but larger indels account for the bulk of lineage-specific sequence in the two genomes.

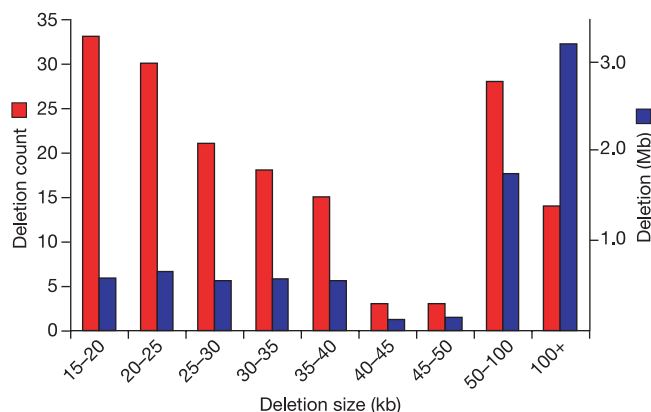


Figure 6 | Length distribution of large indel events (>15 kb), as determined using paired-end sequences from chimpanzee mapped against the human genome. Both the total number of candidate human insertions/chimpanzee deletions (blue) and the number of bases altered (red) are shown.

The retrotransposon SVA and distribution of CpG islands by transposable elements. The third most active element since speciation has been SVA, which created about 1,000 copies in each lineage. SVA is a composite element (~1.5–2.5 kb) consisting of two Alu fragments, a tandem repeat and a region apparently derived from the 3' end of a HERV-K transcript; it is probably mobilized by L1 (refs 79, 80). This element is of particular interest because each copy carries a sequence that satisfies the definition of a CpG island⁸¹ and contains potential transcription factor binding sites; the dispersion of 1,000 SVA copies could therefore be a source of regulatory differences between chimpanzee and human (Supplementary Table S20). At least three human genes contain SVA insertions near their promoters (Supplementary Table S21), one of which has been found to be differentially expressed between the two species^{82,83}, but additional investigations will be required to determine whether the SVA insertion directly caused this difference.

Homologous recombination between interspersed repeats. Human–chimpanzee comparison also makes it possible to study homologous recombination between nearby repeat elements as a source of genomic deletions. We found 612 deletions (totalling 2 Mb) in the human genome that appear to have resulted from recombination between two nearby Alu elements present in the common ancestor; there are 914 such events in the chimpanzee genome. (The events are not biased to (A + T)-rich DNA and thus would not explain the preferential loss of Alu elements in such regions discussed above.) Similarly, we found 26 and 48 instances involving adjacent L1 copies and 8 and 22 instances involving retroviral LTRs in human and chimpanzee, respectively. None of the repeat-mediated deletions removed an orthologous exon of a known human gene in chimpanzee.

The genome comparison allows one to estimate the dependency of homologous recombination on divergence and distance. Homologous recombination seems to occur between quite (>25%) diverged copies (Fig. 8), whereas the number of recombination events (n) varies inversely with the distance (d , in bases) between the copies (as $n \approx 6 \times 10^6 d^{-1.7}$; $r^2 = 0.9$).

Large-scale rearrangements. Finally, we examined the chimpanzee genome sequence for information about large-scale genomic alterations. Cytogenetic studies have shown that human and chimpanzee chromosomes differ by one chromosomal fusion, at least nine pericentric inversions, and in the content of constitutive heterochromatin⁸⁴. Human chromosome 2 resulted from a fusion of two ancestral chromosomes that remained separate in the chimpanzee lineage (chromosomes 2A and 2B in the revised nomenclature¹⁸, formerly chimpanzee chromosomes 12 and 13); the precise fusion point has been mapped and its duplication structure described in detail^{85,86}. In accord with this, alignment of the human and chimpanzee genome sequences shows a break in continuity at this point.

We searched the chimpanzee genome sequence for the precise locations of the 18 breakpoints corresponding to the 9 pericentric inversions (Supplementary Table S22). By mapping paired-end sequences from chimpanzee large insert clones to the human genome, we were able to identify 13 of the breakpoints within the

assembly from discordant end alignments. The positions of five breakpoints (on chromosomes 4, 5 and 12) were tested by fluorescence *in situ* hybridization (FISH) analysis and all were confirmed. Also, the positions of three previously mapped inversion breakpoints (on chromosomes 15 and 18) matched closely those found in the assembly^{87,88}. The paired-end analysis works well in regions of unique sequence, which constitute the bulk of the genome, but is less effective in regions of recent duplication owing to ambiguities in mapping of the paired-end sequences. Beyond the known inversions, we also found suggestive evidence of many additional smaller inversions, as well as older segmental duplications (<98% identity; Supplementary Fig. S6). However, both smaller inversions and more recent segmental duplications will require further investigations.

Gene evolution

We next sought to use the chimpanzee sequence to study the role of natural selection in the evolution of human protein-coding genes. Genome-wide comparisons can shed light on many central issues, including: the magnitude of positive and negative selection; the variation in selection across different lineages, chromosomes, gene families and individual genes; and the complete loss of genes within a lineage.

We began by identifying a set of 13,454 pairs of human and chimpanzee genes with unambiguous 1:1 orthology for which it was possible to generate high-quality sequence alignments covering virtually the entire coding region (Supplementary Information 'Gene evolution' and Table S23). The list contains a large fraction of the entire complement of human genes, although it under-represents gene families that have undergone recent local expansion (such as olfactory receptors and immunoglobulins). To facilitate comparison with the murid lineage, we also compiled a set of 7,043 human, chimpanzee, mouse and rat genes with unambiguous 1:1:1:1 orthology and high-quality sequence alignments (Supplementary Table S24).

Average rates of evolution. To assess the rate of evolution for each gene, we estimated K_A , the number of coding base substitutions that result in amino acid change as a fraction of all such possible sites (the non-synonymous substitution rate). Because the background

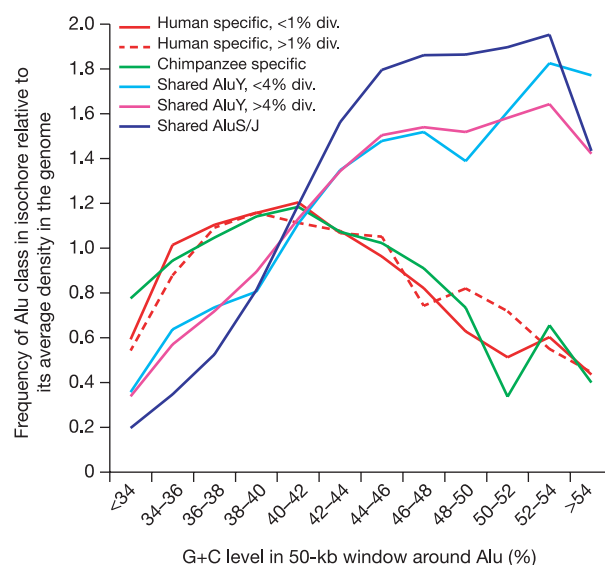


Figure 7 | Correlation of Alu age and distribution by G + C content. Alu elements that inserted after human–chimpanzee divergence are densest in the (G + C)-poor regions of the genome (peaking at 36–40% G + C), whereas older copies, common to both genomes, crowd (G + C)-rich regions. The figure is similar to figure 23 of ref. 24, but the use of chimpanzee allows improved separation of young and old elements, leading to a sharper transition in the pattern.

Table 2 | Transposable element activity in human and chimpanzee lineages

Element	Chimpanzee*	Human*
Alu	2,340 (0.7 Mb)	7,082 (2.1 Mb)
LINE-1	1,979 (>5 Mb)	1,814 (5.0 Mb)
SVA	757 (>1 Mb)	970 (1.3 Mb)
ERV class 1	234 (>1 Mb)†	5 (8 kb)‡
ERV class 2	45 (55 kb)§	77 (130 kb)§
(Micro)satellite	7,054 (4.1 Mb)	11,101 (5.1 Mb)

*Number of lineage-specific insertions (with total size of inserted sequences indicated in brackets) in the aligned parts of the genomes.

†PHERV1 and PHERV2.

‡HERV9.

§Mostly HERV-K.

mutation rate varies across the genome, it is crucial to normalize K_A for comparisons between genes. A striking illustration of this variation is the fact that the mean K_A is 37% higher in the rapidly diverging distal 10 Mb of chromosomes than in the more proximal regions. Classically, the background rate is estimated by K_S , the synonymous substitution rate (coding base substitutions that, because of codon redundancy, do not result in amino acid change). Because a typical gene has only a few synonymous changes between humans and chimpanzees, and not infrequently is zero, we exploited the genome sequence to estimate the local intergenic/intronic substitution rate, K_I , where appropriate. K_A and K_S were also estimated for each lineage separately using mouse and rat as outgroups (Fig. 9).

The K_A/K_S ratio is a classical measure of the overall evolutionary constraint on a gene, where $K_A/K_S \ll 1$ indicates that a substantial proportion of amino acid changes must have been eliminated by purifying selection. Under the assumption that synonymous substitutions are neutral, $K_A/K_S > 1$ implies, but is not a necessary condition for, adaptive or positive selection. The K_A/K_I ratio has the same interpretation. The ratios will sometimes be denoted below by ω with an appropriate subscript (for example, ω_{human}) to indicate the branch of the evolutionary tree under study.

Evolutionary constraint on amino acid sites within the hominid lineage. Overall, human and chimpanzee genes are extremely similar, with the encoded proteins identical in the two species in 29% of cases. The median number of non-synonymous and synonymous substitutions per gene are two and three, respectively. About 5% of the proteins show in-frame indels, but these tend to be small (median = 1 codon)

and to occur in regions of repeated sequence. The close similarity of human and chimpanzee genes necessarily limits the ability to make strong inferences about individual genes, but there is abundant data to study important sets of genes.

The K_A/K_S ratio for the human–chimpanzee lineage (ω_{hominid}) is 0.23. The value is much lower than some recent estimates based on limited sequence data (ranging as high as 0.63 (ref. 7)), but is consistent with an estimate (0.22) from random expressed-sequence-tag (EST) sequencing⁴⁵. Similarly, K_A/K_I was also estimated as 0.23.

Under the assumption that synonymous mutations are selectively neutral, the results imply that 77% of amino acid alterations in hominid genes are sufficiently deleterious as to be eliminated by natural selection. Because synonymous mutations are not entirely neutral (see below), the actual proportion of amino acid alterations with deleterious consequences may be higher. Consistent with previous studies⁸, we find that K_A/K_S of human polymorphisms with frequencies up to 15% is significantly higher than that of human–chimpanzee differences and more common polymorphisms (Table 3), implying that at least 25% of the deleterious amino acid alterations may often attain readily detectable frequencies and thus contribute significantly to the human genetic load.

Evolutionary constraint on synonymous sites within hominid lineage. We next explored the evolutionary constraints on synonymous sites, specifically fourfold degenerate sites. Because such sites have no effect on the encoded protein, they are often considered to be selectively neutral in mammals.

We re-examined this assumption by comparing the divergence at fourfold degenerate sites with the divergence at nearby intronic sites. Although overall divergence rates are very similar at fourfold degenerate and intronic sites, direct comparison is misleading because the former have a higher frequency of the highly mutable CpG dinucleotides (9% compared with 2%). When CpG and non-CpG sites are considered separately, we find that both CpG sites and non-CpG sites show markedly lower divergence in exonic synonymous sites than in introns (~50% and ~30% lower, respectively). This result resolves recent conflicting reports based on limited data sets^{45,89} by showing that such sites are indeed under constraint.

The constraint does not seem to result from selection on the usage of preferred codons, which has been detected in lower organisms⁹⁰ such as bacteria⁹¹, yeast⁹² and flies⁹³. In fact, divergence at fourfold

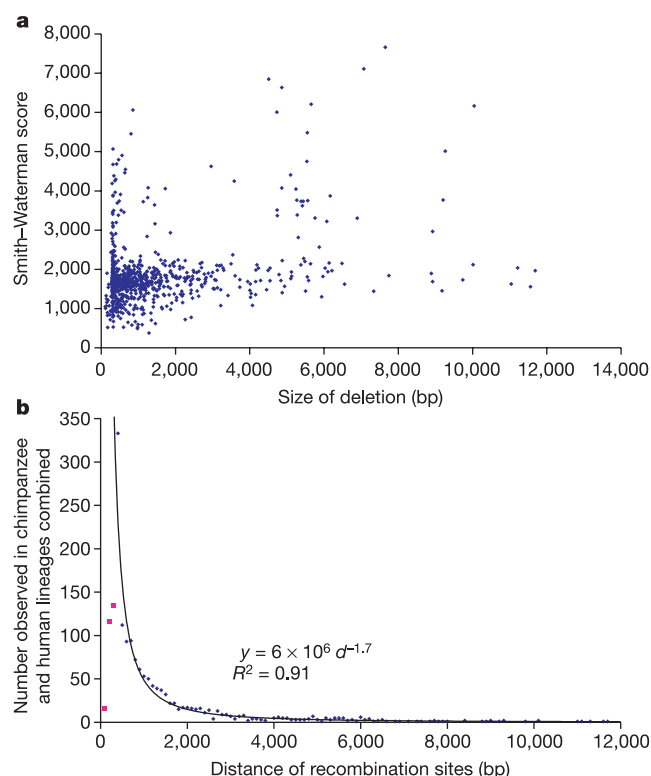


Figure 8 | Dependency of homologous recombination between Alu elements on divergence and distance. **a**, Whereas homologous recombination occurs between quite divergent (Smith–Waterman score $<1,000$), closely spaced copies, more distant recombination seems to favour a better match between the recombining repeats. **b**, The frequency of Alu–Alu-mediated recombination falls markedly as a function of distance between the recombining copies. The first three points (magenta) involve recombination between left or right arms of one Alu inserted into another. The high number of occurrences at a distance of 300–400 nucleotides is due to the preference of integration in the A-rich tail; exclusion of this point does not change the parameters of the equation.

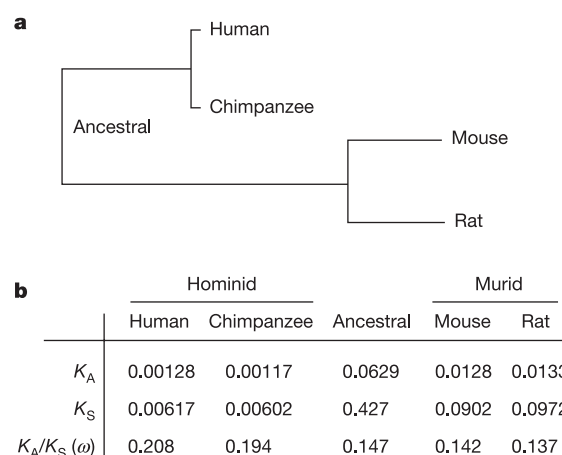


Figure 9 | Human–chimpanzee–mouse–rat tree with branch-specific K_A/K_S (ω) values. **a**, Evolutionary tree. The branch lengths are proportional to the absolute rates of amino acid divergence. **b**, Maximum-likelihood estimates of the rates of evolution in protein-coding genes for humans, chimpanzees, mice and rats. In the text, ω_{hominid} is the K_A/K_S of the combined human and chimpanzee branches and ω_{murid} of the combined mouse and rat branches. The slight difference between ω_{human} and $\omega_{\text{chimpanzee}}$ is not statistically significant; masking of some heterozygous bases in the chimpanzee sequence may contribute to the observed difference (see Supplementary Information ‘Gene evolution’).

degenerate sites increases slightly with codon usage bias (Kendall's $\tau = 0.097$, $P < 10^{-14}$). Alternatively, the observed constraint at synonymous sites might reflect 'background selection'—that is, the indirect effect of purifying selection at amino acid sites causing reduced diversity and thereby reduced divergence at closely linked sites⁴². Given the low rate of recombination in hominid genomes (a 1 kb region experiences only ~ 1 crossover per 100,000 generations or 2 million years), such background selection should extend beyond exons to include nearby intronic sites⁹⁴. However, when the divergence rate is plotted relative to exon–intron boundaries, we find that the rate jumps sharply within a short region of ~ 7 bp at the boundary (Fig. 10). This pattern strongly suggests that the action of purifying selection at synonymous sites is direct rather than indirect, suggesting that other signals, for example those involved in splice site selection, may be embedded in the coding sequence and therefore constrain synonymous sites.

Comparison with murids. An accurate estimate of K_A/K_S makes it possible to study how evolutionary constraint varies across clades. It was predicted more than 30 years ago⁹⁵ that selection against deleterious mutations would depend on population size, with mutations being strongly selected only if they reduce fitness by $s \gg 1/4N$ (where N is effective population size). This would predict that genes would be under stronger purifying selection in murids than hominids, owing to their presumed larger population size. Initial analyses (involving fewer than 50 genes⁹⁶) suggested a strong effect, but the wide variation in estimates of K_A/K_S in hominids^{7,8,97} and murids⁹⁸ has complicated this analysis⁴⁵.

Using the large collection of 7,043 orthologous quartets, we calculated mean K_A/K_S values for the various branches of the four-species evolutionary tree (human, chimpanzee, mouse and rat; Fig. 9). The K_A/K_S ratio for hominids is 0.20. (This is slightly lower than the value of 0.23 obtained with all human–chimpanzee orthologues, probably reflecting slightly greater constraint on the class of proteins with clear orthologues across hominids and murids.)

The K_A/K_S ratio is markedly lower for murids than for hominids ($\omega_{\text{murid}} \approx 0.13$ compared with $\omega_{\text{hominid}} \approx 0.20$) (Fig. 9). This implies that there is an $\sim 35\%$ excess of the amino-acid-changing mutations in the two hominids, relative to the two murids. Excess amino acid divergence may be explained by either increased adaptive evolution or relaxation of evolutionary constraints. As shown in the next section, the latter seems to be the principal explanation.

Relaxed constraints in human evolution. The K_A/K_S ratio can be used to make inferences about the role of positive selection in human evolution^{99,100}. Because alleles under positive selection spread rapidly through a population, they will be found less frequently as common human polymorphisms than as human–chimpanzee differences⁸. Positive selection can thus be detected by comparing the K_A/K_S ratio for common human polymorphisms with the K_A/K_S ratio for

hominid divergence. These ratios have been estimated as $\omega_{\text{polymorphism}} \approx 0.20$ based on an initial collection of common SNPs in human genes and $\omega_{\text{divergence}} \approx 0.34$ based on comparison of human and Old World monkey genes⁸. Thus, the proportion of amino acid changes attributable to positive selection was inferred to be $\sim 35\%$ (ref. 8). This would imply a huge quantitative role for positive selection in human evolution.

With the availability of extensive data for both human polymorphism and human–chimpanzee divergence, we repeated this analysis (using the same set of genes for both estimates). We find that $\omega_{\text{polymorphism}} \approx 0.21$ – 0.23 and $\omega_{\text{divergence}} \approx 0.23$ are statistically indistinguishable (Table 3). Although some of the amino acid substitutions in human and chimpanzee evolution must surely reflect positive selection, the results indicate that the proportion of changes fixed by positive selection seems to be much lower than the previous estimate⁸. (Because the previous results involved comparison to Old World monkeys, it is possible that they reflect strong positive selection earlier in primate evolution; however, we suspect that they reflect the fact that relatively few genes were studied and that different genes were used to study polymorphism and divergence.)

Relaxed negative selection pressures thus primarily explain the excess amino acid divergence in hominid genes relative to murids. Moreover, because both ω_{human} and $\omega_{\text{chimpanzee}}$ are similarly elevated this explanation applies equally to both lineages.

We next sought to study variation in the evolutionary rate of genes within the hominid lineage by searching for unusually high or low levels of constraint for genes and sets of genes.

Rapid evolution in individual genes. We searched for individual genes that have accumulated amino acid substitutions faster than expected given the neutral substitution rate; we considered these genes as potentially being under strong positive selection. A total of 585 of the 13,454 human–chimpanzee orthologues (4.4%) have observed $K_A/K_I > 1$ (see Supplementary Information 'Gene evolution'). However, given the low divergence, the K_A/K_I statistic has large variance. Simulations show that estimates of $K_A/K_I > 1$ would be expected to occur simply by chance in at least 263 cases if purifying selection is allowed to act non-uniformly across genes (Supplementary Fig. S7).

Nonetheless, this set of 585 genes may be enriched for genes that are under positive selection. The most extreme outliers include glycophorin C, which mediates one of the *Plasmodium falciparum* invasion pathways in human erythrocytes¹⁰¹; granulysin, which mediates antimicrobial activity against intracellular pathogens such as *Mycobacterium tuberculosis*¹⁰²; as well as genes that have previously been shown to be undergoing adaptive evolution, such as the protamines and semenogelins involved in reproduction¹⁰³ and the Mas-related gene family involved in nociception¹⁰⁴. With similar

Table 3 | Comparison of K_A/K_S for divergence and human diversity

Substitution type	ΔA	ΔS	K_A/K_S	Per cent excess*	Confidence interval†
Human–chimpanzee divergence	38,773	61,737	0.23	–	–
HapMap (European ancestry)‡					
Rare derived alleles (<15%)	1,614	1,540	0.39	67	[59, 75]
Common alleles	1,199	1,907	0.23	0	[–5, 6]
Frequent derived alleles (>85%)	209	356	0.22	–7	[–19, 7]
HapMap (African ancestry)‡					
Rare derived alleles (<5%)	849	842	0.36	61	[50, 72]
Common alleles	495	803	0.22	–2	[–10, 7]
Frequent derived alleles (>85%)	59	82	0.26	15	[–11, 48]
Affymetrix 120K (multi-ethnic)§					
Rare derived alleles (<15%)	74	82	0.33	44	[14, 80]
Common alleles	77	137	0.21	–11	[–28, 12]
Frequent derived alleles (>85%)	10	15	0.25	6	[–42, 95]

ΔA , Number of observed non-synonymous substitutions. ΔS , Number of observed synonymous substitutions.

* A negative value indicates excess of non-synonymous divergence over polymorphism.

† 95% confidence intervals assuming non-synonymous substitutions are Poisson distributed.

‡ Source: <http://www.hapmap.org> (Public Release no. 13).

§ Source: <http://www.affymetrix.com>.

follow-up studies on candidates from this list, one may be able to draw conclusions about positive selection on other individual genes. In subsequent sections, we examine the rate of divergence for sets of related genes with the aim of detecting subtler signals of accelerated evolution.

Variation in evolutionary rate across physically linked genes. We explored how the rate of evolution varies regionally across the genome. Several studies of mammalian gene evolution have noted that the rate of amino acid substitution shows local clustering, with proteins encoded by nearby genes evolving at correlated rates^{16,105–107}. *Variation across chromosomes.* On the basis of an analysis of ~100 genes¹⁰⁸, it was recently reported that the normalized rate of protein evolution is greater on the nine chromosomes that underwent major structural rearrangement during human evolution (chromosomes 1, 2, 5, 9, 12, 15, 16, 17 and 18); it was suggested that such rearrangements led to reduced gene flow and accelerated adaptive evolution. A subsequent study of a collection of chimpanzee ESTs gave contradictory results^{109,110}. With our larger data set, we re-examined this issue and found no evidence of accelerated evolution on chromosomes with major rearrangements, even if we considered each rearrangement separately (Supplementary Table S25).

Among all hominid chromosomes, the most extreme outlier is chromosome X with a mean K_A/K_I of 0.32. The higher mean seems to reflect a skewed distribution at both high and low values, with the median value (0.17) being more in line with other chromosomes (0.15). The excess of low values may reflect greater purifying selection at some genes, owing to the hemizyosity of chromosome X in males. The excess of high values may reflect increased adaptive selection also resulting from hemizyosity, if a considerable proportion of advantageous alleles are recessive¹¹¹. Interestingly, the higher K_A/K_I value on the X chromosome versus autosomes is largely restricted to genes expressed in testis⁸³.

Variation in local gene clusters. We next searched for genomic neighbourhoods with an unusually high density of rapidly evolving genes. Specifically, we calculated the median K_A/K_I for sliding windows of ten orthologues and identified extreme outliers ($P < 0.001$ compared to random ordering of genes; see Supplementary Information ‘Gene evolution’). A total of 16 such neighbourhoods were found, which greatly exceeds random expectation (Table 4). Repeating the analysis with larger windows (25, 50 and 100 orthologues) did not identify additional rapidly diverging regions.

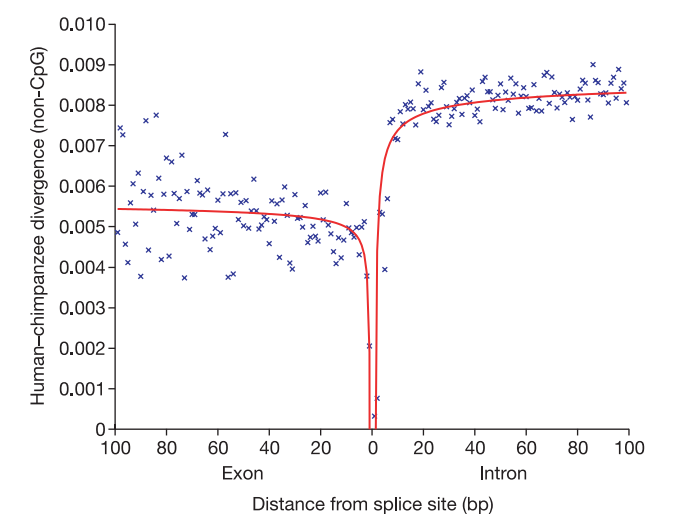


Figure 10 | Purifying selection on synonymous sites. Mean divergence around exon boundaries at non-CpG, exonic, fourfold degenerate sites and intronic sites, relative to the closest mRNA splice junction. The divergence rate at exonic, fourfold degenerate sites is significantly lower than at nearby intronic sites (Mann–Whitney U -test; $P < 10^{-27}$), suggesting that purifying selection limits the rate of synonymous codon substitutions.

In nearly all cases, the regions contain local clusters of phylogenetically and functionally related genes. The rapid diversification of gene families, postulated by ref. 112, can thus be readily discerned even at the relatively close distance of human–chimpanzee divergence. Most of the clusters are associated with functional categories such as host defence and chemosensation (see below). Examples include the epidermal differentiation complex encoding proteins that help form the cornified layer of the skin barrier (Supplementary Fig. S8), the WAP-domain cluster encoding secreted protease inhibitors with antibacterial activity, and the Siglec cluster encoding CD33-related genes. Rapid evolution in these clusters does not seem to be unique to either human or chimpanzee^{113,114}.

Variation in evolutionary rate across functionally related genes. We next studied variation in the evolutionary rate of functional categories of genes, based on the Gene Ontology (GO) classification¹¹⁵.

Rapidly and slowly evolving categories within the hominid lineage. We started by searching for sets of functionally related genes with exceptionally high or low constraint in humans and chimpanzees. For each of the 809 categories with at least 20 genes, K_A/K_S was calculated by concatenating the gene sequences. The category-specific ratios were compared to the average across all orthologues to identify extreme outliers using a metric based on the binomial test (Supplementary Information ‘Gene evolution’ and Supplementary Tables S26–S29). The numbers of observed outliers below a specific threshold (test statistic < 0.001) were then compared to the expected distribution of outliers given randomly permuted annotations.

A total of 98 categories showed elevated K_A/K_S ratios at the specified threshold (Table 5). Only 30 would be expected by chance, indicating that most (but not all) of these categories undergo significantly accelerated evolution relative to the genome-wide average ($P < 10^{-4}$). The rapidly evolving categories within the hominid lineage are primarily related to immunity and host defence, reproduction, and olfaction, which are the same categories known to be undergoing rapid evolution within the broader mammalian lineage, as well as more distantly related species^{15,16,116}. Hominids thus seem to be typical of mammals in this respect (but see below).

A total of 251 categories showed significantly low K_A/K_S ratios (compared with ~32 expected by chance; $P < 10^{-4}$). These include a wide range of processes including intracellular signalling, metabolism, neurogenesis and synaptic transmission, which are evidently under stronger-than-average purifying selection. More generally, genes expressed in the brain show significantly stronger average constraint than genes expressed in other tissues⁸³.

Differences between hominid and murid lineages. Having found gene categories that show substantial variation in absolute evolutionary rate within hominids, we next examined variation in relative rates

Table 4 | Rapidly diverging gene clusters in human and chimpanzee

Location (human)	Cluster	Median K_A/K_I *
1q21	Epidermal differentiation complex	1.46
6p22	Olfactory receptors and HLA-A	0.96
20p11	Cystatins	0.94
19q13	Pregnancy-specific glycoproteins	0.94
17q21	Hair keratins and keratin-associated proteins	0.93
19q13	CD33-related Siglecs	0.90
20q13	WAP domain protease inhibitors	0.90
22q11	Immunoglobulin- λ /breakpoint critical region	0.85
12p13	Taste receptors, type 2	0.81
17q12	Chemokine (C-C motif) ligands	0.81
19q13	Leukocyte-associated immunoglobulin-like receptors	0.80
5q31	Protocadherin- β	0.77
1q32	Complement component 4-binding proteins	0.76
21q22	Keratin-associated proteins and uncharacterized ORFs	0.76
1q23	CD1 antigens	0.72
4q13	Chemokine (C-X-C motif) ligands	0.70

*Maximum median K_A/K_I if the cluster stretched over more than one window of ten genes.

Table 5 | GO categories with the highest divergence rates in hominids

GO categories within 'biological process'	Number of orthologues	Amino acid divergence	K_A/K_S
GO:0007606 sensory perception of chemical stimulus	59	0.018	0.590
GO:0007608 perception of smell	41	0.018	0.521
GO:0006805 xenobiotic metabolism	40	0.013	0.432
GO:0006956 complement activation	22	0.013	0.428
GO:0042035 regulation of cytokine biosynthesis	20	0.011	0.402
GO:0007565 pregnancy	34	0.014	0.384
GO:0007338 fertilization	24	0.010	0.371
GO:0008632 apoptotic programme	36	0.010	0.358
GO:0007283 spermatogenesis	80	0.008	0.354
GO:0000075 cell cycle checkpoint	27	0.006	0.354

Listed are the ten categories in the taxonomy biological process with the highest K_A/K_S ratios, which are not significant solely due to significant subcategories.

between murids and hominids. The K_A/K_S of each of the GO categories are highly correlated between the hominid and murid orthologue pairs, suggesting that the selective pressures acting on particular functional categories have been largely proportional in recent hominid and recent murid evolution (Fig. 11). However, there are several categories with significantly accelerated non-synonymous divergence on each of the lineages, which might represent functions that have undergone lineage-specific positive selection or a lineage-specific relaxation of constraint (Supplementary Information 'Gene evolution' and Supplementary Tables S30–S39).

A total of 59 categories (compared with 11 expected at random, $P < 0.0003$) show evidence of accelerated non-synonymous divergence in the murid lineage. These are dominated by functions and processes related to host defence, such as immune response and lymphocyte activation. Examples include genes encoding interleukins and various T-cell surface antigens (*Cd4*, *Cd8*, *Cd80*). Combined with the recent observation that genes involved in host defence have undergone gene family expansion in murids^{16,17}, this suggests that the immune system has undergone extensive lineage-specific innovation in murids. Additional categories that also show relative acceleration in murids include chromatin-associated proteins and proteins involved in DNA repair. These categories may have similarly undergone stronger adaptive evolution in murids or, alternatively, they may contain fewer sites for mutations with slightly deleterious effects (with the result that the K_A/K_S ratios are less affected by the differences in population size^{96,117}).

Another 58 categories (versus 14 expected at random, $P < 0.0005$) show evidence of accelerated evolution in hominids, with the set dominated by genes encoding proteins involved in transport (for example, ion transport), synaptic transmission, spermatogenesis and perception of sound (Table 6). Notably, some outliers include genes with brain-related functions, compatible with a recent finding¹¹⁸. Potential positive selection on spermatogenesis genes in the hominids was also recently noted¹¹⁹. However, as above, it is possible that these categories could have more sites for slightly deleterious mutations and thus be more affected by population size differences. Sequence information from more species and from individuals

within species will be necessary to distinguish between the possible explanations.

Differences between the human and chimpanzee lineage. One of the most interesting questions is perhaps whether certain categories have undergone accelerated evolution in humans relative to chimpanzees, because such genes might underlie unique aspects of human evolution.

As was done for hominids and murids above, we compared non-synonymous divergence for each category to search for relative acceleration in either lineage (Fig. 12). Seven categories show signs of accelerated evolution on the human lineage relative to chimpanzee, but this is only slightly more than the four expected at random ($P < 0.22$). Intriguingly, the single strongest outlier is 'transcription factor activity', with the 348 human genes studied having accumulated 47% more amino acid changes than their chimpanzee orthologues. Genes with accelerated divergence in human include homeotic, forkhead and other transcription factors that have key roles in early development. However, given the small number of changes involved, additional data will be required to confirm this trend. There was no excess of accelerated categories on the chimpanzee lineage.

We also compared human genes with and without disease associations, including mental retardation, for differences in mutation rate when compared to chimpanzee. Briefly, no significant differences were observed in either the background mutation rate or in the ratio of human-specific changes to chimpanzee-specific amino acid changes (see Supplementary Information 'Gene evolution' and Supplementary Tables S40 and S41).

We thus find minimal evidence of acceleration unique to either the human or chimpanzee lineage across broad functional categories. This is not simply due to general lack of power resulting from the small number of changes since the divergence of human and chimpanzee, because one can detect acceleration of categories in either hominid relative to either murid. For example, 29 accelerated categories versus 9 expected at random ($P < 0.02$) can be detected on the human lineage, and 40 categories versus 11 expected at random ($P < 0.007$) on the chimpanzee lineage, relative to mouse. But the

Table 6 | GO categories with accelerated divergence rates in hominids relative to murids

GO categories within 'biological process'	Number of orthologues	Amino acid divergence in hominids	Amino acid divergence in murids	K_A/K_S in hominids	K_A/K_S in murids
GO:0007283 spermatogenesis	43	0.0075	0.054	0.323	0.188
GO:0006869 lipid transport	22	0.0081	0.051	0.306	0.120
GO:0006865 amino acid transport	24	0.0058	0.033	0.218	0.084
GO:0015698 inorganic anion transport	29	0.0061	0.027	0.195	0.072
GO:0006486 protein amino acid glycosylation	50	0.0056	0.040	0.166	0.100
GO:0019932 second-messenger-mediated signalling	58	0.0049	0.036	0.159	0.083
GO:0007605 perception of sound	28	0.0052	0.033	0.158	0.085
GO:0016051 carbohydrate biosynthesis	27	0.0047	0.028	0.147	0.067
GO:0007268 synaptic transmission	93	0.0040	0.025	0.126	0.069
GO:0006813 potassium ion transport	65	0.0035	0.022	0.113	0.056

Listed are the ten categories in the taxonomy biological process with the strongest evidence for accelerated evolution in hominids relative to murids, which are not significant solely due to significant subcategories.

outliers are largely the same for both human and chimpanzee, indicating that the fraction of amino acid mutations that have contributed to human- and chimpanzee-specific patterns of evolution must be small relative to the fraction that have contributed to a common hominid and, to a large extent, mammalian pattern of evolution.

It was recently reported¹⁰ that several functional categories are enriched for genes with evidence of positive selection in the human lineage or the chimpanzee lineage, and that these categories are largely different between the two lineages. These results and ours differ in ways that will require further investigation. With the potential exception of some developmental regulators, the categories that ref. 10 reported as showing the strongest enrichment of positive selection in one lineage (including cell adhesion, ion transport and perception of sound) are among those that we show as having accelerated divergence in both human and chimpanzee. This suggests that positive selection and relaxation of constraints may be correlated, or alternatively, that the results of ref. 10 may be enriched for false positives in categories that have experienced particularly strong relaxation of constraints in the hominids. Data from additional primates, as well as advances in analytical methods, will be necessary to distinguish between these alternatives. At present, strong evidence of positive selection unique to the human lineage is thus limited to a handful of genes¹²⁰.

Our analysis above largely omitted genes belonging to large gene families, because gene family expansion makes it difficult to define 1:1:1 orthologues across hominids and murids. One of the largest such families, the olfactory receptors, is known to be undergoing rapid divergence in primates. Directed study of these genes in the draft assembly has suggested that more than 100 functional human olfactory receptors are likely to be under no evolutionary constraint¹²¹. Our analysis also omitted the majority of very recently duplicated genes owing to their lower coverage in the current chimpanzee assembly. However, recent human-specific duplications can be readily identified from the finished human genome sequence, and have previously been shown to be highly enriched for the same categories found to have high absolute rates of evolution in 1:1 orthologues here; that is, olfaction, immunity and reproduction²³.

Gene disruptions in human and chimpanzee. Whereas most genes have undergone only subtle substitutions in their amino acid sequence, a few dozen have suffered more marked changes. We found a total of 53 known or predicted human genes that are either deleted entirely (36) or partially (17) in chimpanzee (Supplementary

Table S42). We have so far tested and confirmed 15 of these cases by polymerase chain reaction (PCR) or Southern blotting. An additional eight genes have sustained large deletions (>15 kb) entirely within an intron. Some genes may have been missed in this count owing to limitations of the draft genome sequence. In addition, some genes may have suffered chain termination mutations or altered reading frames in chimpanzee, but accurate identification of these will require higher-quality sequence. The sensitivity of the reciprocal analysis of genes disrupted in human is currently limited by the small number of independently predicted gene models for the chimpanzee. Some of the gene disruptions may be related to interesting biological differences between the species, as discussed below.

Genetic basis for human- and chimpanzee-specific biology. Given the substantial number of neutral mutations, only a small subset of the observed gene differences is likely to be responsible for the key phenotypic changes in morphology, physiology and behavioural complexity between humans and chimpanzees. Determining which differences are in this evolutionarily important subset and inferring their functional consequences will require additional types of evidence, including information from clinical observations and model systems¹²². We describe some novel examples of genetic changes for which plausible functional or physiological consequences can be suggested.

Apoptosis. Mouse and human are known to differ with respect to an important mediator of apoptosis, caspase-12 (refs 123–125). The protein triggers apoptosis in response to perturbed calcium homeostasis in mice, but humans seem to lack this activity owing to several mutations in the orthologous gene that together affect the protein produced by all known splice forms; the mutations include a premature stop codon and a disruption of the SHG box required for enzymatic activity of caspases. By contrast, the chimpanzee gene encodes an intact open reading frame and SHG box, indicating that the functional loss occurred in the human lineage. Intriguingly, loss-of-function mutations in mice confer increased resistance to amyloid-induced neuronal apoptosis without causing obvious developmental or behavioural defects¹²⁶. The loss of function in humans may contribute to the human-specific pathology of Alzheimer's disease, which involves amyloid-induced neurotoxicity and deranged calcium homeostasis.

Inflammatory response. Human and chimpanzee show a notable difference with respect to important mediators of immune and inflammatory responses. Three genes (*IL1F7*, *IL1F8* and *ICEBERG*)

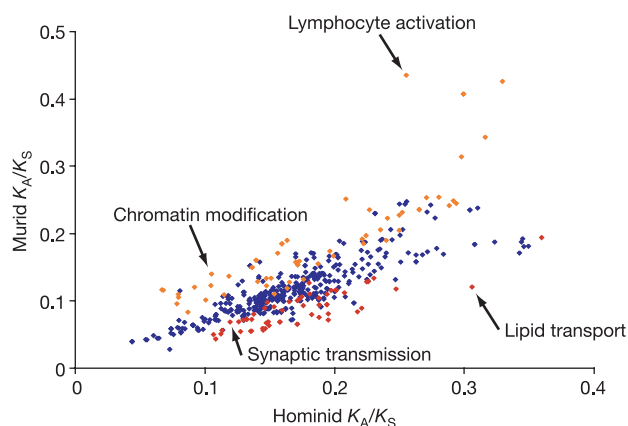


Figure 11 | Hominid and murid K_A/K_S (ω) in GO categories with more than 20 analysed genes. GO categories with putatively accelerated (test statistic <0.001; see Methods) non-synonymous divergence on the hominid lineages (red) and on the murid lineages (orange) are highlighted. Owing to the hierarchical nature of GO, the categories do not all represent independent data points. A non-redundant list of significant categories is provided in Table 8 and a complete list in Supplementary Table S30.

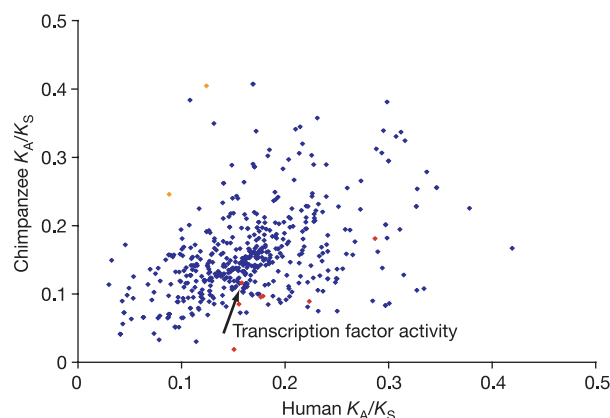


Figure 12 | Human and chimpanzee K_A/K_S (ω) in GO categories with more than 20 analysed genes. GO categories with putatively accelerated (test statistic <0.001; see Methods) non-synonymous divergence on the human lineage (red) and on the chimpanzee lineage (orange) are highlighted. The variance of these estimates is larger than that seen in the hominid–murid comparison owing to the small number of lineage-specific substitutions. Owing to the hierarchical nature of the GO ontology, the categories do not all represent independent data points. A complete list of categories is provided in Supplementary Table S30.

that act in a common pathway involving the caspase-1 gene all appear to be deleted in chimpanzee. *ICEBERG* is thought to repress caspase-1-mediated generation of pro-inflammatory *IL1* cytokines, and its absence in chimpanzee may point to species-specific modulation of the interferon- γ - and lipopolysaccharide-induced inflammatory response¹²⁷.

Parasite resistance. Similarly, we found that two members of the primate-specific *APOL* gene cluster (*APOL1* and *APOL4*) have been deleted from the chimpanzee genome. The *APOL1* protein is associated with the high-density lipoprotein fraction in serum and has recently been proposed to be the lytic factor responsible for resistance to certain subspecies of *Trypanosoma brucei*, the parasite that causes human sleeping sickness and the veterinary disease nagana¹²⁸. The loss of the *APOL1* gene in chimpanzees could thus explain the observation that human, gorilla and baboon possess the trypanosome lytic factor, whereas the chimpanzee does not¹²⁹.

Sialic acid biology related proteins. Sialic acids are cell-surface sugars that mediate many biological functions¹³⁰. Of 54 genes involved in sialic acid biology, 47 were suitable for analysis. We confirmed and extended findings on several that have undergone human-specific changes, including disruptions, deletions and domain-specific functional changes^{113,131,132}. Human- and chimpanzee-specific changes were also found in otherwise evolutionarily conserved sialyl motifs in four sialyl transferases (*ST6GAL1*, *ST6GALNAC3*, *ST6GALNAC4* and *ST8SIA2*), suggesting changes in donor and/or acceptor binding¹³⁰. Lineage-specific changes were found in a complement factor H (*HF1*) sialic acid binding domain associated with human disease¹³³. Human *SIGLEC11* has undergone gene conversion with a nearby pseudogene, correlating with acquisition of human-specific brain expression and altered binding properties¹³⁴.

Human disease alleles. We next sought to identify putative functional differences between the species by searching for instances in which a human disease-causing allele appears to be the wild-type allele in the chimpanzee. Starting from 12,164 catalogued disease variants in 1,384 human genes, we identified 16 cases in which the altered sequence in a disease allele matched the chimpanzee sequence, and had plausible support in the literature (Table 7; see also Supplementary Table S43). Upon re-sequencing in seven chimpanzees, 15 cases were confirmed homozygous in all individuals, whereas one (*PON1* I102V) appears to be a shared polymorphism (Supplementary Table S44).

Six cases represent *de novo* human mutations associated with simple mendelian disorders. Similar cases have also been found in comparisons of more distantly related mammals¹³⁵, as well as

between insects¹³⁶, and have been interpreted as a consequence of a relatively high rate of compensatory mutations. If compensatory mutations are more likely to be fixed by positive selection than by neutral drift¹³⁶, then the variants identified here might point towards adaptive differences between humans and chimpanzees. For example, the ancestral Thr 29 allele of cationic trypsinogen (*PRSS1*) causes autosomal dominant pancreatitis in humans¹³⁷, suggesting that the human-specific Asn 29 allele may represent a digestion-related molecular adaptation¹³⁸.

The remaining ten cases represent common human polymorphisms that have been reported to be associated with complex traits, including coronary artery disease and diabetes mellitus. In all of these cases we confirmed that the disease-associated allele in humans is indeed the ancestral allele by showing that it is carried not only by chimpanzee but also by outgroups such as the macaque. These ancestral alleles may thus have become human-specific risk factors due to changes in human physiology or environment, and the polymorphisms may represent ongoing adaptations. For example, *PPARG* Pro 12 is the wild-type allele in chimpanzee but has been clearly associated with increased risk of type 2 diabetes in human¹³⁹. It is tempting to speculate that this allele may represent an ancestral 'thrifty' genotype¹⁴⁰.

The current results must be interpreted with caution, because few complex disease associations have been firmly established. The fact that the human disease allele is the wild-type allele in chimpanzee may actually indicate that some of the putative associations are spurious and not causal. However, this approach can be expected to become increasingly fruitful as the quality and completeness of the disease mutation databases improve.

Human population genetics

The chimpanzee has a special role in informing studies of human population genetics, a field that is undergoing rapid expansion and acquiring new relevance to human medical genetics¹⁴¹. The chimpanzee sequence allows recognition of those human alleles that represent the ancestral state and the derived state. It also allows estimates of local mutation rates, which serve as an important baseline in searching for signs of natural selection.

Ancestral and derived alleles. Of ~7.2 million SNPs mapped to the human genome in the current public database, we could assign the alleles as ancestral or derived in 80% of the cases according to which allele agrees with the chimpanzee genome sequence¹⁴² (see Supplementary Information 'Human population genetics'). For the remaining cases, no assignment could be made because of the following: the orthologous chimpanzee base differed from both human alleles (1.2%); was polymorphic in the chimpanzee sequences obtained (0.4%); or could not be reliably identified with the current draft sequence of the chimpanzee (18.8%), with many of these occurring in repeated or segmentally duplicated sequence. The first two cases arise presumably because a second mutation occurred in the chimpanzee lineage. It should be possible to resolve most of these cases by examining a close outgroup such as gorilla or orang-utan.

Mutations in the chimpanzee may also lead to the erroneous assignment of human alleles as derived alleles. This error rate can be estimated as the probability of a second mutation resulting in the chimpanzee sequence matching the derived allele (see Supplementary Information 'Human population genetics'). The estimated error rate for typical SNPs is 0.5%, owing to the low nucleotide substitution rate. The exceptions are those SNPs for which the human alleles are CpG and TpG and the chimpanzee sequence is TpG. For these, a non-negligible fraction may have arisen by two independent deamination events within an ancestral CpG dinucleotide, which are well-known mutational hotspots⁵¹ (also see above). Human SNPs in a CpG context for which the orthologous chimpanzee sequence is TpG account for 12% of the total, and have an estimated error rate of 9.8%. Across all SNPs, the average error rate, ϵ , is thus estimated to be ~1.6%.

We compared the distribution of allele frequencies for ancestral

Table 7 | Candidate human disease variants found in chimpanzee

Gene	Variant*	Disease association	Ancestral†	Frequency‡
<i>AIRE</i>	P252L ¹⁵⁹	Autoimmune syndrome	Unresolved	0
<i>MKKS</i>	R518H ¹⁶⁰	Bardet-Biedl syndrome	Wild type	0
<i>MLH1</i>	A441T ¹⁶¹	Colorectal cancer	Wild type	0
<i>MYOC</i>	Q48H ¹⁶²	Glaucoma	Wild type	0
<i>OTC</i>	T125M ¹⁶³	Hyperammonaemia	Wild type	0
<i>PRSS1</i>	N29T ¹³⁷	Pancreatitis	Disease	0
<i>ABCA1</i>	I883M ¹⁶⁴	Coronary artery disease	Unresolved	0.136
<i>APOE</i>	C130R ¹⁶⁵	Coronary artery disease and Alzheimer's disease	Disease	0.15
<i>DIO2</i>	T92A ¹⁶⁶	Insulin resistance	Disease	0.35
<i>ENPP1</i>	K121Q ¹⁶⁷	Insulin resistance	Disease	0.17
<i>GSTP1</i>	I105V ¹⁶⁸	Oral cancer	Disease	0.348
<i>PON1S</i>	I102V ¹⁶⁹	Prostate cancer	Wild type	0.016
<i>PON1</i>	Q192R ¹⁷⁰	Coronary artery disease	Disease	0.3
<i>PPARG</i>	A12P ¹³⁹	Type 2 diabetes	Disease	0.85
<i>SLC2A2</i>	T110I ¹⁷¹	Type 2 diabetes	Disease	0.12
<i>UCP1</i>	A64T ¹⁷²	Waist-to-hip ratio	Disease	0.12

* This takes the following format: benign variant, codon number, disease/chimpanzee variant.

† Ancestral variant as inferred from closest available primate outgroups (Supplementary Information).

‡ Frequency of the disease allele in human study population.

§ Polymorphic in chimpanzee.

and derived alleles using a database of allele frequencies for $\sim 120,000$ SNPs (see Supplementary Information 'Human population genetics'). As expected, ancestral alleles tend to have much higher frequencies than derived alleles (Supplementary Fig. S9). Nonetheless, a significant proportion of derived alleles have high frequencies: 9.1% of derived alleles have frequency $\geq 80\%$.

An elegant result in population genetics states that, for a randomly interbreeding population of constant size, the probability that an allele is ancestral is equal to its frequency¹⁴³. We explored the extent to which this simple theoretical expectation fits the human population. We tabulated the proportion $p_a(x)$ of ancestral alleles for various frequencies of x and compared this with the prediction $p_a(x) = x$ (Fig. 13).

The data lie near the predicted line, but the observed slope (0.83) is substantially less than 1. One explanation for this deviation is that some ancestral alleles are incorrectly assigned (an error rate of ε would artificially decrease the slope by a factor of $1 - 2\varepsilon$). However, with ε estimated to be only 1.6%, errors can only explain a small part of the deviation. The most likely explanation is the presence of bottlenecks during human history, which tend to flatten the distribution of allele frequencies. Theoretical calculations indicate that a recent bottleneck would decrease the slope by a factor of $(1 - b)$, where b is the inbreeding coefficient induced by the bottleneck (see Supplementary Information 'Human population genetics' and Supplementary Fig. S10). This suggests that measurements of the slope in different human groups may shed light on population-specific bottlenecks. Consistent with this, preliminary analyses of allele frequencies in several regions for SNPs obtained by systematic uniform sampling indicate that the slope is significantly lower than 1 in European and Asian samples and close to 1 in an African sample (see Supplementary Information 'Human population genetics' and Supplementary Fig. S11).

Signatures of strong selective sweeps in recent human history. The pattern of human genetic variation holds substantial information about selection events that have shaped our species. Strong positive selection creates the distinctive signature of a 'selective sweep', whereby a rare allele rapidly rises to fixation and carries the haplotype on which it occurs to high frequency (the 'hitchhiking' effect). The surrounding region should show two distinctive signatures: a significant reduction of overall diversity, and an excess of derived alleles with high frequency in the population owing to hitchhiking of

derived alleles on the selected haplotype (see Supplementary Information 'Human population genetics'). The pattern might be detectable for up to 250,000 years after a selective sweep has ended¹⁴⁴. Notably, the chimpanzee genome provides crucial baseline information required for accurate assessment of both signatures.

The size of the interval affected by a selective sweep is expected to scale roughly with s , the selective advantage due to the mutation. Simulations can be used to study the distribution of the interval size (see Supplementary Information 'Human population genetics'). With $s = 1\%$, the interval over which heterozygosity falls by 50% has a modal size of 600 kb and a probability of greater than 10% of exceeding 1 Mb.

We undertook an initial scan for large regions (>1 Mb) with the two signatures suggestive of strong selective sweeps in recent human history. We began by identifying regions in which the observed human diversity rate was much lower than the expectation based on the observed divergence rate with chimpanzee. The human diversity rate was measured as the number of occurrences from a database of 1.92 million SNPs identified by shotgun sequencing in a panel of African-American individuals (see Supplementary Information 'Genome sequencing and assembly'). The comparison with the chimpanzee eliminates regions in which low diversity simply reflects a low mutation rate in the region. Regions were identified based on a simple statistical procedure (see Supplementary Information 'Human population genetics'). Six genomic regions stand out as clear outliers that show significantly reduced diversity relative to divergence (Table 8; see also Supplementary Fig. S12).

We next tested whether these six regions show a high proportion of SNPs with high-frequency derived alleles (defined here as alleles with frequency $\geq 80\%$). Within each region, we focused on the 1-Mb interval with the greatest discrepancy between diversity and divergence and compared it to 1-Mb regions throughout the genome. For the database of 120,000 SNPs with allele frequencies discussed above, the typical 1-Mb region in the human genome contains ~ 40 SNPs, and the proportion p_h of SNPs with high-frequency derived alleles is $\sim 9.1\%$. All six regions identified by our scan for reduced diversity have a higher than average fraction of high-frequency derived alleles; all six fall within the top 10% genome-wide and three fall within the top 1%. Although this is not definitive evidence for any particular region, the joint probability of all six regions randomly scoring in the top 10% is 10^{-6} . The results indicate that the six regions are candidates for strong selective sweeps during the past 250,000 years¹⁴⁴. The regions differ notably with respect to gene content, ranging from one containing 57 annotated genes (chromosome 22) to another with no annotated genes whatsoever (chromosome 4). We have no evidence to implicate any individual functional element as a target of recent selection at this point, but the regions contain a number of interesting candidates for follow-up studies. Intriguingly, the chromosome 4 gene desert, which flanks a proto-cadherin gene and is conserved across vertebrates¹⁵, has been implicated in two independent studies as being associated with obesity^{145,146}.

In addition to the six regions, one further genomic region deserves mention: an interval of 7.6 Mb on chromosome 7q (see Supplementary Information 'Human population genetics'). The interval contains several regions with high scores in the diversity-divergence analysis (including the seventh highest score overall) as well as in the proportion of high-frequency derived alleles. The region contains the *FOXP2* and *CFTR* genes. The former has been the subject of much interest as a possible target for selection during human evolution¹⁴⁷ and the latter as a target of selection in European populations¹⁴⁸.

Convincing proof of past selection will require careful analysis of the precise pattern of genetic variation in the region and the identification of a likely target of selection. Nonetheless, our findings suggest that the approach outlined here may help to unlock some of the secrets of recent human evolution through a combination of within-species and cross-species comparison.

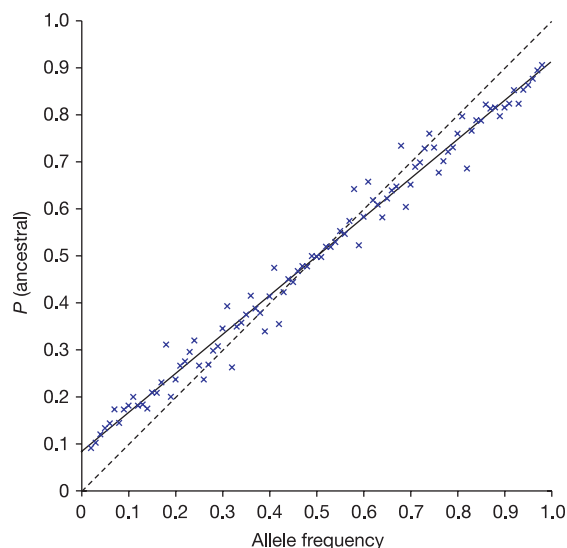


Figure 13 | The observed fraction of ancestral alleles in 1% bins of observed frequency. The solid line shows the regression ($b = 0.83$). The dotted line shows the theoretical relationship $p_a(x) = x$. Note that because each variant yields a derived and an ancestral allele, the data are necessarily symmetrical about 0.5.

Table 8 | Human regions with strongest signal of selection based on diversity relative to divergence

Chromosome	Start (Mb)	End (Mb)	Regression log-score	Skew P-value	Genes
1	48.58	52.58	103.3	0.071	Fourteen known genes from ELAVL4 to GPX7
2	144.35	148.47	84.8	0.074	ARHGAP15 (partial), GTDC1 and ZFXH1B
22	36.15	40.22	81.8	0.00022	Fifty-seven known genes from CARD10 to PMM1
12	84.69	89.01	80.9	0.031	Ten known genes from PAMC1 to ATP2B1
8	34.91	37.54	76.9	0.00032	UNCSD and FKSG2
4	32.42	35.62	55.9	0.00067	No known genes or Ensembl predictions

Discussion

Our knowledge of the human genome is greatly advanced by the availability of a second hominid genome. Some questions can be directly answered by comparing the human and chimpanzee sequences, including estimates of regional mutation rates and average selective constraints on gene classes. Other questions can be addressed in conjunction with other large data sets, such as issues in human population genetics for which the chimpanzee genome provides crucial controls. For still other questions, the chimpanzee genome simply provides a starting point for further investigation.

The hardest such question is: what makes us human? The challenge lies in the fact that most evolutionary change is due to neutral drift. Adaptive changes comprise only a small minority of the total genetic variation between two species. As a result, the extent of phenotypic variation between organisms is not strictly related to the degree of sequence variation. For example, gross phenotypic variation between human and chimpanzee is much greater than between the mouse species *Mus musculus* and *Mus spretus*, although the sequence difference in the two cases is similar. On the other hand, dogs show considerable phenotypic variation despite having little overall sequence variation (~0.15%). Genomic comparison markedly narrows the search for the functionally important differences between species, but specific biological insights will be needed to sift the still-large list of candidates to separate adaptive changes from neutral background.

Our comparative analysis suggests that the patterns of molecular evolution in the hominids are typical of a broader class of mammals in many ways, but distinctive in certain respects. As with the murids, the most rapidly evolving gene families are those involved in reproduction and host defence. In contrast to the murids, however, hominids appear to experience substantially weaker negative selection; this probably reflects their smaller population size. Consequently, hominids accumulate deleterious mutations that would be eliminated by purifying selection in murids. This may be both an advantage and a disadvantage. Although decreased purifying selection may tend to erode overall fitness, it may also allow hominids to 'explore' larger regions of the fitness landscape and thereby achieve evolutionary adaptations that can only be reached by passing through intermediate states of inferior fitness^{149,150}.

Although the analyses presented here focus on protein-coding sequences, the chimpanzee genome sequence also allows systematic analysis of the recent evolution of gene regulatory elements for the first time. Initial analysis of both gene expression patterns and promoter regions suggest that their overall patterns of evolution closely mirror that of protein-coding regions. In an accompanying paper⁸³, we show that the rates of change in gene expression among different tissues in human and chimpanzee correlate with the nucleotide divergence in the putative proximal promoters and even more interestingly with the average level of constraint on proteins in the same tissues. Another study¹⁵¹ has similarly used the chimpanzee sequence described here to show that gene promoter regions are also evolving under markedly less constraint in hominids than in murids.

The draft chimpanzee sequence here is sufficient for initial analyses, but it is still imperfect and incomplete. Definitive studies of gene and genome evolution—including pseudogene formation, gene family expansion and segmental duplication—will require high-

quality finished sequence. In this regard, we note that efforts are already underway to construct a BAC-based physical map and to increase the shotgun sequence coverage to approximately sixfold redundancy. The added coverage alone will not affect the analysis greatly, but plans are in place to produce finished sequence for difficult to sequence and important segments of the genome.

Our close biological relatedness to chimpanzees not only allows unique insights into human biology, it also creates ethical obligations. Although the genome sequence was acquired without harm to chimpanzees, the availability of the sequence may increase pressure to use chimpanzees in experimentation. We strongly oppose reducing the protection of chimpanzees and instead advocate the policy positions suggested by an accompanying paper¹⁵². Furthermore, the existence of chimpanzees and other great apes in their native habitats is increasingly threatened by human civilization. More effective policies are urgently needed to protect them in the wild. We hope that elaborating how few differences separate our species will broaden recognition of our duty to these extraordinary primates that stand as our siblings in the family of life.

METHODS

Sequencing and assembly. Approximately 22.5 million sequence reads were derived from both ends of inserts (paired end reads) from 4-, 10-, 40- and 180-kb clones, all prepared from primary blood lymphocyte DNA. Genomic resources available from the source animal include a lymphoid cell line (S006006) and genomic DNA (NS06006) at Coriell Cell Repositories (<http://locus.umdj.edu/ccr/>), as well as a BAC library (CHORI-251)¹⁵³ (see also Supplementary Information 'Genome sequencing and assembly').

Genome alignment. BLASTZ¹⁵⁴ was used to align non-repetitive chimpanzee regions against repeat-masked human sequence. BLAT¹⁵⁵ was subsequently used to align the more repetitive regions. The combined alignments were chained¹⁵⁶ and only best reciprocal alignments were retained for further analysis.

Insertions and deletions. Small insertion/deletion (indel) events (<15 kb) were parsed directly from the BLASTZ genome alignment by counting the number and size of alignment gaps between bases within the same contig. Sites of large-scale indels (>15 kb) were detected from discordant placements of paired sequence reads against the human assembly. Size thresholds were obtained from both human fosmid alignments on human sequence (40 ± 2.58 kb) and chimpanzee plasmid alignments against human chromosome 21 (4.5 ± 1.84 kb). Indels were inferred by two or more pairs surpassing these thresholds by more than two standard deviations and the absence of sequence data within the discordancy.

Gene annotation. A total of 19,277 human RefSeq transcripts¹⁵⁷, representing 16,045 distinct genes, were indirectly aligned to the chimpanzee sequence via the genome alignment. After removing low-quality sequences and likely alignment artefacts, an initial catalogue containing 13,454 distinct 1:1 human–chimpanzee orthologues was created for the analyses described here. A subset of 7,043 of these genes with unambiguous mouse and rat orthologues were realigned using Clustal W¹⁵⁸ for the lineage-specific analyses. Updated gene catalogues can be obtained from <http://www.ensembl.org>.

Rates of divergence. Nucleotide divergence rates were estimated using baseml with the REV model. Non-CpG rates were estimated from all sites that did not overlap a CG dinucleotide in either human or chimpanzee. K_A and K_S were estimated jointly for each orthologue using codeml with the F3x4 codon frequency model and no additional constraints, except for the comparison of divergent and polymorphic substitutions where K_A/K_S for both was estimated as $(\Delta A/N_A)/(\Delta S/N_S)$, with N_S/N_A , the ratio of synonymous to non-synonymous sites, estimated as 0.36 from the orthologue alignments. Unless otherwise specified, K_A/K_S for a set of genes was calculated by summing the number of substitutions and the number of sites to obtain K_A and K_S for the concatenated set before taking

the ratio. Hominid and murid pairwise rates were estimated independently from codons aligned across all four species. Human and chimpanzee lineage-specific K_A and K_S were estimated on an unrooted tree with both mouse and rat included. Lineage-specific rates were also estimated by parsimony, with essentially identical results (see Supplementary Information). K_1 was estimated from all interspersed repeats within 250 kb of the mid-point of each gene.

Accelerated evolution in GO categories. The binomial probability of observing X or more non-synonymous substitutions, given a total of $X + Y$ substitutions and the expected proportion x from all orthologues, was calculated by summing substitutions across the orthologues in each GO category. For the absolute rate test, Y = the number of synonymous substitutions in orthologues in the same category. For the relative rate tests, Y = the number of non-synonymous substitutions on the opposite lineage. Note that this binomial probability is simply a metric designed to identify potentially accelerated categories, it is not a P -value that can be used to reject directly the null hypothesis of no acceleration in that particular category. For each test, the observed number of categories with a binomial probability less than 0.001 was compared to the expected distribution of such outliers by repeating the procedure 10,000 times on randomly permuted GO annotations. The significance of the number of observed outliers n was estimated as the proportion of random trials yielding n or more outliers.

Detection of selective sweeps. The observed number of human SNPs, u_i , human bases, m_i , human–chimpanzee substitutions, v_i , and chimpanzee bases, n_i , within each set of non-overlapping 1-Mb windows along the human genome were used to generate two random numbers, x_i (adjusted human diversity) and y_i (adjusted human–chimpanzee divergence), from the two beta-distributions:

$$x_i \approx \text{Beta}(u_i + a, m_i - u_i + b)$$

$$y_i \approx \text{Beta}(v_i + c, n_i - v_i + d)$$

where $a = 1$, $b = 1,000$, $c = 1$ and $d = 100$. These numbers were then fit to a linear regression:

$$x|y \approx N(\alpha_0 + \alpha_1 y, \beta^2)$$

A P -value for each window was calculated for each window based on (x_i, y_i) and the regression line. This was repeated 100 times and the average of the P -values taken as the P -value for diversity given divergence in each window. Overlapping windows with $P < 0.1$ containing at least one window of $P < 0.05$ were coalesced and scored as the sum of their $-\log(p)$ scores.

Received 21 March; accepted 20 July 2005.

- Darwin, C. *The Descent of Man, and Selection in Relation to Sex* (D Appleton and Company, New York, 1871).
- Huxley, T. H. *Evidence as to Man's Place in Nature* (Williams and Norgate, London, 1863).
- Goodman, M. The genomic record of humankind's evolutionary roots. *Am. J. Hum. Genet.* **64**, 31–39 (1999).
- Goodall, J. Tool-using and aimed throwing in a community of free-living chimpanzees. *Nature* **201**, 1264–1266 (1964).
- Whiten, A. *et al.* Cultures in chimpanzees. *Nature* **399**, 682–685 (1999).
- Olson, M. V. & Varki, A. Sequencing the chimpanzee genome: insights into human evolution and disease. *Nature Rev. Genet.* **4**, 20–28 (2003).
- Eyre-Walker, A. & Keightley, P. D. High genomic deleterious mutation rates in hominids. *Nature* **397**, 344–347 (1999).
- Fay, J. C., Wyckoff, G. J. & Wu, C. I. Positive and negative selection on the human genome. *Genetics* **158**, 1227–1234 (2001).
- King, M. C. & Wilson, A. C. Evolution at two levels in humans and chimpanzees. *Science* **188**, 107–116 (1975).
- Clark, A. G. *et al.* Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios. *Science* **302**, 1960–1963 (2003).
- Hellmann, I. *et al.* Selection on human genes as revealed by comparisons to chimpanzee cDNA. *Genome Res.* **13**, 831–837 (2003).
- Ebersberger, I., Metzler, D., Schwarz, C. & Paabo, S. Genomewide comparison of DNA sequences between humans and chimpanzees. *Am. J. Hum. Genet.* **70**, 1490–1497 (2002).
- Watanabe, H. *et al.* DNA sequence and comparative analysis of chimpanzee chromosome 22. *Nature* **429**, 382–388 (2004).
- Jaillon, O. *et al.* Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature* **431**, 946–957 (2004).
- Hillier, L. W. *et al.* Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432**, 695–716 (2004).
- Mouse Genome Sequencing Consortium, Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562 (2002).
- Rat Genome Sequencing Project Consortium. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* **428**, 493–521 (2004).
- McConkey, E. H. Orthologous numbering of great ape and human chromosomes is essential for comparative genomics. *Cytogenet. Genome Res.* **105**, 157–158 (2004).
- Sanger, F., Coulson, A. R., Hong, G. F., Hill, D. F. & Petersen, G. B. Nucleotide sequence of bacteriophage lambda DNA. *J. Mol. Biol.* **162**, 729–773 (1982).
- Myers, G. Whole-genome DNA sequencing. *Comput. Sci. Eng.* **1**, 33–43 (1999).
- Huang, X., Wang, J., Aluru, S., Yang, S. P. & Hillier, L. PCAP: a whole-genome assembly program. *Genome Res.* **13**, 2164–2170 (2003).
- Jaffe, D. B. *et al.* Whole-genome sequence assembly for mammalian genomes: Arachne 2. *Genome Res.* **13**, 91–96 (2003).
- International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931–945 (2004).
- International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–920 (2001).
- Ewing, B., Hillier, L., Wendl, M. C. & Green, P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* **8**, 175–185 (1998).
- She, X. *et al.* Shotgun sequence assembly and recent segmental duplications within the human genome. *Nature* **431**, 927–930 (2004).
- Cheng, Z. *et al.* A genome-wide comparison of recent chimpanzee and human segmental duplications. *Nature* doi:10.1038/nature04000 (this issue).
- Fischer, A., Wiebe, V., Paabo, S. & Przeworski, M. Evidence for a complex demographic history of chimpanzees. *Mol. Biol. Evol.* **21**, 799–808 (2004).
- Yu, N. *et al.* Low nucleotide diversity in chimpanzees and bonobos. *Genetics* **164**, 1511–1518 (2003).
- Kaessmann, H., Wiebe, V., Weiss, G. & Paabo, S. Great ape DNA sequences reveal a reduced diversity and an expansion in humans. *Nature Genet.* **27**, 155–156 (2001).
- Kitano, T., Schwarz, C., Nickel, B. & Paabo, S. Gene diversity patterns at 10 X-chromosomal loci in humans and chimpanzees. *Mol. Biol. Evol.* **20**, 1281–1289 (2003).
- The International SNP Map Working Group. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* **409**, 928–933 (2001).
- Chen, F. C. & Li, W. H. Genomic divergences between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees. *Am. J. Hum. Genet.* **68**, 444–456 (2001).
- Fujiyama, A. *et al.* Construction and analysis of a human-chimpanzee comparative clone map. *Science* **295**, 131–134 (2002).
- Hardison, R. C. *et al.* Covariation in frequencies of substitution, deletion, transposition, and recombination during eutherian evolution. *Genome Res.* **13**, 13–26 (2003).
- Webster, M. T., Smith, N. G., Lercher, M. J. & Ellegren, H. Gene expression, synteny, and local similarity in human noncoding mutation rates. *Mol. Biol. Evol.* **21**, 1820–1830 (2004).
- Rosenberg, H. F. & Feldmann, M. W. *The Relationship Between Coalescence Times and Population Divergence Times* (Oxford Univ. Press, Oxford, 2002).
- Vignaud, P. *et al.* Geology and palaeontology of the Upper Miocene Toros-Menalla hominid locality, Chad. *Nature* **418**, 152–155 (2002).
- Wall, J. D. Estimating ancestral population sizes and divergence times. *Genetics* **163**, 395–404 (2003).
- Reich, D. E. *et al.* Human genome sequence variation and the influence of gene history, mutation and recombination. *Nature Genet.* **32**, 135–142 (2002).
- Maynard Smith, J. M. & Haigh, J. The hitch-hiking effect of a favourable gene. *Genet. Res.* **23**, 23–35 (1974).
- Hudson, R. R. & Kaplan, N. L. Deleterious background selection with recombination. *Genetics* **141**, 1605–1617 (1995).
- Charlesworth, B. The effect of background selection against deleterious mutations on weakly selected, linked variants. *Genet. Res.* **63**, 213–227 (1994).
- Birky, C. W. Jr & Walsh, J. B. Effects of linkage on rates of molecular evolution. *Proc. Natl Acad. Sci. USA* **85**, 6414–6418 (1988).
- Hellmann, I., Ebersberger, I., Ptak, S. E., Paabo, S. & Przeworski, M. A neutral explanation for the correlation of diversity with recombination rates in humans. *Am. J. Hum. Genet.* **72**, 1527–1535 (2003).
- Lercher, M. J. & Hurst, L. D. Human SNP variability and mutation rate are higher in regions of high recombination. *Trends Genet.* **18**, 337–340 (2002).
- Hellmann, I. *et al.* Why do human diversity levels vary at a megabase scale? *Genome Res.* **15**, 1222–1231 (2005).
- Li, W. H., Yi, S. & Makova, K. Male-driven evolution. *Curr. Opin. Genet. Dev.* **12**, 650–656 (2002).
- Bohossian, H. B., Skaletsky, H. & Page, D. C. Unexpectedly similar rates of nucleotide substitution found in male and female hominids. *Nature* **406**, 622–625 (2000).
- Makova, K. D. & Li, W. H. Strong male-driven evolution of DNA sequences in humans and apes. *Nature* **416**, 624–626 (2002).
- Hwang, D. G. & Green, P. Bayesian Markov chain Monte Carlo sequence analysis reveals varying neutral substitution patterns in mammalian evolution. *Proc. Natl Acad. Sci. USA* **101**, 13994–14001 (2004).
- Taylor, J., Tyekucheva, S., Zody, M., Ciaromonte, F. & Makova, K. D. Strong and weak male mutation bias at different sites in the primate genomes: Insights from the human-chimpanzee comparison. *Mol. Biol. Evol.* (submitted).
- Bulmer, M., Wolfe, K. H. & Sharp, P. M. Synonymous nucleotide substitution

- rates in mammalian genes: implications for the molecular clock and the relationship of mammalian orders. *Proc. Natl Acad. Sci. USA* **88**, 5974–5978 (1991).
54. Ehrlich, M., Zhang, X. Y. & Inamdar, N. M. Spontaneous deamination of cytosine and 5-methylcytosine residues in DNA and replacement of 5-methylcytosine residues with cytosine residues. *Mutat. Res.* **238**, 277–286 (1990).
 55. Craig, J. M. & Bickmore, W. A. Chromosome bands—flavours to savour. *Bioessays* **15**, 349–354 (1993).
 56. Holmquist, G. P. Chromosome bands, their chromatin flavors, and their functional features. *Am. J. Hum. Genet.* **51**, 17–37 (1992).
 57. Ellegren, H., Smith, N. G. & Webster, M. T. Mutation rate variation in the mammalian genome. *Curr. Opin. Genet. Dev.* **13**, 562–568 (2003).
 58. Cooper, G. M., Brudno, M., Green, E. D., Batzoglou, S. & Sidow, A. Quantitative estimates of sequence divergence for comparative analyses of mammalian genomes. *Genome Res.* **13**, 813–820 (2003).
 59. Cooper, G. M. *et al.* Characterization of evolutionary rates and constraints in three mammalian genomes. *Genome Res.* **14**, 539–548 (2004).
 60. Yang, S. *et al.* Patterns of insertions and their covariation with substitutions in the rat, mouse, and human genomes. *Genome Res.* **14**, 517–527 (2004).
 61. Birdsall, J. A. Integrating genomics, bioinformatics, and classical genetics to study the effects of recombination on genome evolution. *Mol. Biol. Evol.* **19**, 1181–1197 (2002).
 62. Jensen-Seaman, M. I. *et al.* Comparative recombination rates in the rat, mouse, and human genomes. *Genome Res.* **14**, 528–538 (2004).
 63. Fortna, A. *et al.* Lineage-specific gene duplication and loss in human and great ape evolution. *PLoS Biol.* **2**, E207 (2004).
 64. Britten, R. J. Divergence between samples of chimpanzee and human DNA sequences is 5%, counting indels. *Proc. Natl Acad. Sci. USA* **99**, 13633–13635 (2002).
 65. Frazer, K. A. *et al.* Genomic DNA insertions and deletions occur frequently between humans and nonhuman primates. *Genome Res.* **13**, 341–346 (2003).
 66. Locke, D. P. *et al.* Large-scale variation among human and great ape genomes determined by array comparative genomic hybridization. *Genome Res.* **13**, 347–357 (2003).
 67. Liu, G. *et al.* Analysis of primate genomic variation reveals a repeat-driven expansion of the human genome. *Genome Res.* **13**, 358–368 (2003).
 68. Yohn, C. T. *et al.* Lineage-specific expansions of retroviral insertions within the genomes of African great apes but not humans and orangutans. *PLoS Biol.* **3**, 1–11 (2005).
 69. Hedges, D. J. *et al.* Differential alu mobilization and polymorphism among the human and chimpanzee lineages. *Genome Res.* **14**, 1068–1075 (2004).
 70. Smit, A. F. Interspersed repeats and other mementos of transposable elements in mammalian genomes. *Curr. Opin. Genet. Dev.* **9**, 657–663 (1999).
 71. Dewannieux, M., Esnault, C. & Heidmann, T. LINE-mediated retrotransposition of marked Alu sequences. *Nature Genet.* **35**, 41–48 (2003).
 72. Mathews, L. M., Chi, S. Y., Greenberg, N., Ovchinnikov, I. & Swergold, G. D. Large differences between LINE-1 amplification rates in the human and chimpanzee lineages. *Am. J. Hum. Genet.* **72**, 739–748 (2003).
 73. Pickeral, O. K., Makalowski, W., Boguski, M. S. & Boeke, J. D. Frequent human genomic DNA transduction driven by LINE-1 retrotransposition. *Genome Res.* **10**, 411–415 (2000).
 74. Goodier, J. L., Ostertag, E. M. & Kazazian, H. H. Jr Transduction of 3'-flanking sequences is common in L1 retrotransposition. *Hum. Mol. Genet.* **9**, 653–657 (2000).
 75. Zhang, Z., Harrison, P. M., Liu, Y. & Gerstein, M. Millions of years of evolution preserved: a comprehensive catalog of the processed pseudogenes in the human genome. *Genome Res.* **13**, 2541–2558 (2003).
 76. Torrents, D., Suyama, M., Zdobnov, E. & Bork, P. A genome-wide survey of human pseudogenes. *Genome Res.* **13**, 2559–2567 (2003).
 77. Esnault, C., Maestre, J. & Heidmann, T. Human LINE retrotransposons generate processed pseudogenes. *Nature Genet.* **24**, 363–367 (2000).
 78. Zhang, Z., Harrison, P. & Gerstein, M. Identification and analysis of over 2000 ribosomal protein pseudogenes in the human genome. *Genome Res.* **12**, 1466–1482 (2002).
 79. Ostertag, E. M., Goodier, J. L., Zhang, Y. & Kazazian, H. H. Jr SVA elements are nonautonomous retrotransposons that cause disease in humans. *Am. J. Hum. Genet.* **73**, 1444–1451 (2003).
 80. Shen, L. *et al.* Structure and genetics of the partially duplicated gene RP located immediately upstream of the complement C4A and the C4B genes in the HLA class III region. Molecular cloning, exon-intron structure, composite retroposon, and breakpoint of gene duplication. *J. Biol. Chem.* **269**, 8466–8476 (1994).
 81. Takai, D. & Jones, P. A. Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proc. Natl Acad. Sci. USA* **99**, 3740–3745 (2002).
 82. Enard, W. *et al.* Intra- and interspecific variation in primate gene expression patterns. *Science* **296**, 340–343 (2002).
 83. Khaitovich, P. *et al.* Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. *Science* (in the press).
 84. Yunis, J. J., Sawyer, J. R. & Dunham, K. The striking resemblance of high-resolution G-banded chromosomes of man and chimpanzee. *Science* **208**, 1145–1148 (1980).
 85. Fan, Y., Linardopoulou, E., Friedman, C., Williams, E. & Trask, B. J. Genomic structure and evolution of the ancestral chromosome fusion site in 2q13-2q14.1 and paralogous regions on other human chromosomes. *Genome Res.* **12**, 1651–1662 (2002).
 86. Fan, Y., Newman, T., Linardopoulou, E. & Trask, B. J. Gene content and function of the ancestral chromosome fusion site in human chromosome 2q13-2q14.1 and paralogous regions. *Genome Res.* **12**, 1663–1672 (2002).
 87. Locke, D. P. *et al.* Refinement of a chimpanzee pericentric inversion breakpoint to a segmental duplication cluster. *Genome Biol.* **4**, R50 (2003).
 88. Dennehey, B. K., Gutches, D. G., McConkey, E. H. & Krauter, K. S. Inversion, duplication, and changes in gene context are associated with human chromosome 18 evolution. *Genomics* **83**, 493–501 (2004).
 89. Subramanian, S. & Kumar, S. Neutral substitutions occur at a faster rate in exons than in noncoding DNA in primate genomes. *Genome Res.* **13**, 838–844 (2003).
 90. Duret, L. Detecting genomic features under weak selective pressure: the example of codon usage in animals and plants. *Bioinformatics* **18** (suppl. 2), S91 (2002).
 91. Sharp, P. M. & Li, W. H. Codon usage in regulatory genes in *Escherichia coli* does not reflect selection for 'rare' codons. *Nucleic Acids Res.* **14**, 7737–7749 (1986).
 92. Sharp, P. M., Averof, M., Lloyd, A. T., Matassi, G. & Peden, J. F. DNA sequence evolution: the sounds of silence. *Phil. Trans. R. Soc. Lond. B* **349**, 241–247 (1995).
 93. Moriyama, E. N. & Powell, J. R. Synonymous substitution rates in *Drosophila*: mitochondrial versus nuclear genes. *J. Mol. Evol.* **45**, 378–391 (1997).
 94. McVean, G. A. *et al.* The fine-scale structure of recombination rate variation in the human genome. *Science* **304**, 581–584 (2004).
 95. Ohta, T. Slightly deleterious mutant substitutions during evolution. *Nature* **246**, 96–98 (1973).
 96. Ohta, T. Synonymous and nonsynonymous substitutions in mammalian genes and the nearly neutral theory. *J. Mol. Evol.* **40**, 56–63 (1995).
 97. Eyre-Walker, A., Keightley, P. D., Smith, N. G. & Gaffney, D. Quantifying the slightly deleterious mutation model of molecular evolution. *Mol. Biol. Evol.* **19**, 2142–2149 (2002).
 98. Makalowski, W. & Boguski, M. S. Synonymous and nonsynonymous substitution distances are correlated in mouse and rat genes. *J. Mol. Evol.* **47**, 119–121 (1998).
 99. McDonald, J. H. & Kreitman, M. Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature* **351**, 652–654 (1991).
 100. Sawyer, S. A. & Hartl, D. L. Population genetics of polymorphism and divergence. *Genetics* **132**, 1161–1176 (1992).
 101. Maier, A. G. *et al.* *Plasmodium falciparum* erythrocyte invasion through glycophorin C and selection for Gerbich negativity in human populations. *Nature Med.* **9**, 87–92 (2003).
 102. Stenger, S. *et al.* An antimicrobial activity of cytolytic T cells mediated by granulysin. *Science* **282**, 121–125 (1998).
 103. Swanson, W. J. & Vacquier, V. D. The rapid evolution of reproductive proteins. *Nature Rev. Genet.* **3**, 137–144 (2002).
 104. Choi, S. S. & Lahn, B. T. Adaptive evolution of MRG, a neuron-specific gene family implicated in nociception. *Genome Res.* **13**, 2252–2259 (2003).
 105. Hardison, R. C. *et al.* Global predictions and tests of erythroid regulatory regions. *Cold Spring Harb. Symp. Quant. Biol.* **68**, 335–344 (2003).
 106. Lercher, M. J., Chamary, J. V. & Hurst, L. D. Genomic regionalization in rates of evolution is not explained by clustering of genes of comparable expression profile. *Genome Res.* **14**, 1002–1013 (2004).
 107. Williams, E. J. & Hurst, L. D. The proteins of linked genes evolve at similar rates. *Nature* **407**, 900–903 (2000).
 108. Navarro, A. & Barton, N. H. Chromosomal speciation and molecular divergence—accelerated evolution in rearranged chromosomes. *Science* **300**, 321–324 (2003).
 109. Zhang, J., Wang, X. & Podlaha, O. Testing the chromosomal speciation hypothesis for humans and chimpanzees. *Genome Res.* **14**, 845–851 (2004).
 110. Lu, J., Li, W. H. & Wu, C. I. Comment on "Chromosomal speciation and molecular divergence—accelerated evolution in rearranged chromosomes". *Science* **302**, 988 (2003).
 111. Charlesworth, B., Coyne, J. A. & Orr, H. A. Meiotic drive and unisexual hybrid sterility: a comment. *Genetics* **133**, 421–432 (1993).
 112. Ohno, S. *Evolution by Gene Duplication* (Springer, New York, 1970).
 113. Angata, T., Margulies, E. H., Green, E. D. & Varki, A. Large-scale sequencing of the CD33-related Siglec gene cluster in five mammalian species reveals rapid evolution by multiple mechanisms. *Proc. Natl Acad. Sci. USA* **101**, 13251–13256 (2004).
 114. Teumer, J. & Green, H. Divergent evolution of part of the involucrin gene in the hominoids: unique intragenic duplications in the gorilla and human. *Proc. Natl Acad. Sci. USA* **86**, 1283–1286 (1989).
 115. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genet.* **25**, 25–29 (2000).
 116. Yang, Z. & Bielawski, J. P. Statistical methods for detecting molecular adaptation. *Trends Ecol. Evol.* **15**, 496–503 (2000).
 117. Weinreich, D. M. The rates of molecular evolution in rodent and primate mitochondrial DNA. *J. Mol. Evol.* **52**, 40–50 (2001).
 118. Dorus, S. *et al.* Accelerated evolution of nervous system genes in the origin of *Homo sapiens*. *Cell* **119**, 1027–1040 (2004).

119. Neilsen, R. *et al.* A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biol.* **3**, e170 (2005).
120. Vallender, E. J. & Lahn, B. T. Positive selection on the human genome. *Hum. Mol. Genet.* **13** (suppl. 2), R245–R254 (2004).
121. Gilad, Y., Man, O. & Glusman, G. A comparison of the human and chimpanzee olfactory receptor gene repertoires. *Genome Res.* **15**, 224–230 (2005).
122. Enard, W. & Paabo, S. Comparative primate genomics. *Annu. Rev. Genomics Hum. Genet.* **5**, 351–378 (2004).
123. Saleh, M. *et al.* Differential modulation of endotoxin responsiveness by human caspase-12 polymorphisms. *Nature* **429**, 75–79 (2004).
124. Fischer, H., Koenig, U., Eckhart, L. & Tschachler, E. Human caspase 12 has acquired deleterious mutations. *Biochem. Biophys. Res. Commun.* **293**, 722–726 (2002).
125. Puente, X. S., Sanchez, L. M., Overall, C. M. & Lopez-Otin, C. Human and mouse proteases: a comparative genomic approach. *Nature Rev. Genet.* **4**, 544–558 (2003).
126. Nakagawa, T. *et al.* Caspase-12 mediates endoplasmic-reticulum-specific apoptosis and cytotoxicity by amyloid- β . *Nature* **403**, 98–103 (2000).
127. Humke, E. W., Shriver, S. K., Starovasnik, M. A., Fairbrother, W. J. & Dixit, V. M. ICEBERG: a novel inhibitor of interleukin-1 β generation. *Cell* **103**, 99–111 (2000).
128. Vanhamme, L. *et al.* Apolipoprotein L-I is the trypanosome lytic factor of human serum. *Nature* **422**, 83–87 (2003).
129. Seed, J. R., Sechelski, J. B. & Loomis, M. R. A survey for a trypanocidal factor in primate sera. *J. Protozool.* **37**, 393–400 (1990).
130. Angata, T. & Varki, A. Chemical diversity in the sialic acids and related α -keto acids: an evolutionary perspective. *Chem. Rev.* **102**, 439–469 (2002).
131. Varki, A. How to make an ape brain. *Nature Genet.* **36**, 1034–1036 (2004).
132. Sonnenburg, J. L., Altheide, T. K. & Varki, A. A uniquely human consequence of domain-specific functional adaptation in a sialic acid-binding receptor. *Glycobiology* **14**, 339–346 (2004).
133. Pangburn, M. K. Host recognition and target differentiation by factor H, a regulator of the alternative pathway of complement. *Immunopharmacology* **49**, 149–157 (2000).
134. Hayakawa, T. *et al.* Human-specific gene in microglia. *Science* (in the press).
135. Kondrashov, A. S., Sunyaev, S. & Kondrashov, F. A. Dobzhansky-Muller incompatibilities in protein evolution. *Proc. Natl Acad. Sci. USA* **99**, 14878–14883 (2002).
136. Kulathinal, R. J., Bettencourt, B. R. & Hartl, D. L. Compensated deleterious mutations in insect genomes. *Science* **306**, 1553–1554 (2004).
137. Pfutzer, R. *et al.* Novel cationic trypsinogen (PRSS1) N29T and R122C mutations cause autosomal dominant hereditary pancreatitis. *Gut* **50**, 271–272 (2002).
138. Chen, J. M., Montier, T. & Ferec, C. Molecular pathology and evolutionary and physiological implications of pancreatitis-associated cationic trypsinogen mutations. *Hum. Genet.* **109**, 245–252 (2001).
139. Altshuler, D. *et al.* The common PPAR γ Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nature Genet.* **26**, 76–80 (2000).
140. Neel, J. V. Diabetes mellitus: a “thrifty” genotype rendered detrimental by “progress”? *Am. J. Hum. Genet.* **14**, 353–362 (1962).
141. The International HapMap Consortium. The International HapMap Project. *Nature* **426**, 789–796 (2003).
142. Hacia, J. G. *et al.* Determination of ancestral alleles for human single-nucleotide polymorphisms using high-density oligonucleotide arrays. *Nature Genet.* **22**, 164–167 (1999).
143. Watterson, G. A. & Guess, H. A. Is the most frequent allele the oldest? *Theor. Popul. Biol.* **11**, 141–160 (1977).
144. Przeworski, M. The signature of positive selection at randomly chosen loci. *Genetics* **160**, 1179–1189 (2002).
145. Stone, S. *et al.* A major predisposition locus for severe obesity, at 4p15-p14. *Am. J. Hum. Genet.* **70**, 1459–1468 (2002).
146. Arya, R. *et al.* Evidence of a novel quantitative-trait locus for obesity on chromosome 4p in Mexican Americans. *Am. J. Hum. Genet.* **74**, 272–282 (2004).
147. Enard, W. *et al.* Molecular evolution of *FOXP2*, a gene involved in speech and language. *Nature* **418**, 869–872 (2002).
148. Schroeder, S. A., Gaughan, D. M. & Swift, M. Protection against bronchial asthma by *CFTR* Δ F508 mutation: a heterozygote advantage in cystic fibrosis. *Nature Med.* **1**, 703–705 (1995).
149. Ohta, T. Evolution by nearly-neutral mutations. *Genetica* **102–103**, 83–90 (1998).
150. Hayakawa, T., Altheide, T. K. & Varki, A. Genetic basis of human brain evolution: accelerating along the primate speedway. *Dev. Cell* **8**, 2–4 (2005).
151. Keightley, P. D., Lercher, M. J. & Eyre-Walker, A. Evidence for widespread degradation of gene control regions in hominid genomes. *PLoS Biol.* **3**, e42 (2005).
152. Gagneux, P., Moore, J. J. & Varki, A. The ethics of research on great apes. *Nature* **437**, 27–29 (2005).
153. Osoegawa, K. *et al.* An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* **52**, 1–8 (1998).
154. Schwartz, S. *et al.* Human-mouse alignments with BLASTZ. *Genome Res.* **13**, 103–107 (2003).
155. Kent, W. J. BLAT—the BLAST-like alignment tool. *Genome Res.* **12**, 656–664 (2002).
156. Kent, W. J., Baertsch, R., Hinrichs, A., Miller, W. & Haussler, D. Evolution’s cauldron: duplication, deletion, and rearrangement in the mouse and human genomes. *Proc. Natl Acad. Sci. USA* **100**, 11484–11489 (2003).
157. Pruitt, K. D., Tatusova, T. & Maglott, D. R. NCBI Reference Sequence project: update and current status. *Nucleic Acids Res.* **31**, 34–37 (2003).
158. Higgins, D. G., Thompson, J. D. & Gibson, T. J. Using CLUSTAL for multiple sequence alignments. *Methods Enzymol.* **266**, 383–402 (1996).
159. Meloni, A. *et al.* Delineation of the molecular defects in the AIRE gene in autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy patients from Southern Italy. *J. Clin. Endocrinol. Metab.* **87**, 841–846 (2002).
160. Beales, P. L. *et al.* Genetic and mutational analyses of a large multiethnic Bardet-Biedl cohort reveal a minor involvement of *BBS6* and delineate the critical intervals of other loci. *Am. J. Hum. Genet.* **68**, 606–616 (2001).
161. Cunningham, J. M. *et al.* The frequency of hereditary defective mismatch repair in a prospective series of unselected colorectal carcinomas. *Am. J. Hum. Genet.* **69**, 780–790 (2001).
162. Mukhopadhyay, A. *et al.* Mutations in MYOC gene of Indian primary open angle glaucoma patients. *Mol. Vis.* **8**, 442–448 (2002).
163. Tuchman, M., Jaleel, N., Morizono, H., Sheehy, L. & Lynch, M. G. Mutations and polymorphisms in the human ornithine transcarbamylase gene. *Hum. Mutat.* **19**, 93–107 (2002).
164. Clee, S. M. *et al.* Common genetic variation in *ABCA1* is associated with altered lipoprotein levels and a modified risk for coronary artery disease. *Circulation* **103**, 1198–1205 (2001).
165. Fullerton, S. M. *et al.* Apolipoprotein E variation at the sequence haplotype level: implications for the origin and maintenance of a major human polymorphism. *Am. J. Hum. Genet.* **67**, 881–900 (2000).
166. Mentuccia, D. *et al.* Association between a novel variant of the human type 2 deiodinase gene Thr92Ala and insulin resistance: evidence of interaction with the Trp64Arg variant of the β -3-adrenergic receptor. *Diabetes* **51**, 880–883 (2002).
167. Pizzuti, A. *et al.* A polymorphism (K121Q) of the human glycoprotein PC-1 gene coding region is strongly associated with insulin resistance. *Diabetes* **48**, 1881–1884 (1999).
168. Katoh, T. *et al.* Human glutathione S-transferase P1 polymorphism and susceptibility to smoking related epithelial cancer; oral, lung, gastric, colorectal and urothelial cancer. *Pharmacogenetics* **9**, 165–169 (1999).
169. Marchesani, M. *et al.* New paraoxonase 1 polymorphism I102V and the risk of prostate cancer in Finnish men. *J. Natl Cancer Inst.* **95**, 812–818 (2003).
170. Humbert, R. *et al.* The molecular basis of the human serum paraoxonase activity polymorphism. *Nature Genet.* **3**, 73–76 (1993).
171. Barroso, I. *et al.* Candidate gene association study in type 2 diabetes indicates a role for genes involved in β -cell function as well as insulin action. *PLoS Biol.* **1**, E20 (2003).
172. Herrmann, S. M. *et al.* Uncoupling protein 1 and 3 polymorphisms are associated with waist-to-hip ratio. *J. Mol. Med.* **81**, 327–332 (2003).
173. Kong, A. *et al.* A high-resolution recombination map of the human genome. *Nature Genet.* **31**, 241–247 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Generation of the *Pan troglodytes* sequence at Washington University School of Medicine’s Genome Sequencing Center and the Broad Institute of MIT and Harvard was supported by grants from the National Human Genome Research Institute (NHGRI). We would like to thank the entire staff of both of those institutions. For work from other groups, we acknowledge the support of the European Molecular Biology Laboratory, Ministerio de Educacion y Ciencia (Spain), Howard Hughes Medical Institute, NHGRI, National Institutes of Health and National Science Foundation. Resources for exploring the sequence and annotation data are available on browser displays available at UCSC (<http://genome.ucsc.edu>), Ensembl (<http://www.ensembl.org>) and the NCBI (<http://www.ncbi.nlm.nih.gov>). We thank L. Gaffney for graphical help.

Author Contributions The last three authors co-directed the work.

Author Information This *Pan troglodytes* whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the project accessions ARACHNE, AADA01000000 and PCAP, AACZ01000000. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.H.V. (waterston@gs.washington.edu) E.S.L. (lander@broad.mit.edu) or R.K.W. (rwilson@watson.wustl.edu).

The Chimpanzee Sequencing and Analysis Consortium Tarjei S. Mikkelsen^{1,2}, LaDeana W. Hillier³, Evan E. Eichler⁴, Michael C. Zody¹, David B. Jaffe¹, Shiaw-Pyng Yang³, Wolfgang Enard⁵, Ines Hellmann⁵, Kerstin Lindblad-Toh¹, Tasha K. Altheide⁶, Nicoletta Archidiacono⁷, Peer Bork^{8,9}, Jonathan Butler¹, Jean L. Chang¹, Ze Cheng⁴, Asif T. Chinwalla³, Pieter deJong¹⁰, Kimberley D. Delehaunty³, Catrina C. Fronick³, Lucinda L. Fulton³, Yoav Gilad¹¹, Gustavo Glusman¹², Sante Gnerre¹, Tina A. Graves³, Toshiyuki Hayakawa⁶, Karen E. Hayden¹³, Xiaohu Huang¹⁴, Hongkai Ji¹⁵, W. James Kent¹⁶, Mary-Claire King⁴, Edward J. Kulbokas III¹, Ming K. Lee⁴, Ge Liu¹³, Carlos Lopez-Otin¹⁷, Kateryna D. Makova¹⁸, Orna Man¹⁹, Elaine R. Mardis³, Evan Mauceli¹, Tracie L. Miner³, William E. Nash³, Joanne O. Nelson³, Svante Pääbo⁵, Nick J. Patterson¹, Craig S. Pohl¹³, Katherine S. Pollard¹⁶, Kay Prüfer⁵, Xose S. Puente¹⁷, David Reich^{1,20}, Mariano Rocchi¹⁷, Kate Rosenbloom¹⁶, Maryellen Ruvolo²¹, Daniel J. Richter¹, Stephen F. Schaffner¹, Arian F. A. Smit¹², Scott M. Smith³, Mikita Suyama⁸, James Taylor¹⁸, David Torrents⁸, Eray Tuzun⁴, Ajit Varki⁶, Gloria Velasco¹⁷, Mario Ventura⁷, John W. Wallis³, Michael C. Wendt¹³, Richard K. Wilson³, Eric S. Lander^{1,22,23,24} & Robert H. Waterston⁴

Affiliations for participants: ¹Broad Institute of MIT and Harvard, 320 Charles Street, Cambridge, Massachusetts 02141, USA. ²Division of Health Sciences and Technology, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA. ³Genome Sequencing Center, Washington University School of Medicine, Campus Box 8501, 4444 Forest Park Avenue, St Louis, Missouri 63108, USA. ⁴Genome Sciences, University of Washington School of Medicine, 1705 NE Pacific Street, Seattle, Washington 98195, USA. ⁵Max Planck Institute of Evolutionary Anthropology, Deutscher Platz 6, D-04103 Leipzig, Germany. ⁶University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093, USA. ⁷Department of Genetics and Microbiology, University of Bari, 70126 Bari, Italy. ⁸EMBL, Meyerhofstrasse 1, Heidelberg D-69117, Germany. ⁹Max Delbrück Center for Molecular Medicine (MDC), Robert-Rössle-Strasse 10, D-13125 Berlin, Germany. ¹⁰Children's Hospital Oakland Research Institute, 747 52nd Street, Oakland, California 94609, USA. ¹¹Department of Genetics, Yale University School of Medicine, 333 Cedar Street, New Haven, Connecticut 06520, USA. ¹²Institute for Systems Biology, 1441 North 34th Street, Seattle, Washington 98103, USA. ¹³Department of Genetics, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, Ohio 44106, USA. ¹⁴Department of Computer Science, Iowa State University, 226 Atanasoff Hall, Ames, Iowa 50011, USA. ¹⁵Department of Statistics, Harvard University, 1 Oxford Street, Cambridge, Massachusetts 02138, USA. ¹⁶University of California, Santa Cruz, Center for Biomolecular Science and Engineering, 1156 High Street, Santa Cruz, California 95064, USA. ¹⁷Departamento de Bioquímica y Biología Molecular, Instituto Universitario de Oncología del Principado de Asturias, Universidad de Oviedo, C/Fernando Bongera s/n, 33006 Oviedo, Spain. ¹⁸The Pennsylvania State University, Center for Comparative Genomics and Bioinformatics and Department of Biology, University Park, Pennsylvania 16802, USA. ¹⁹Department of Structural Biology, Weizmann Institute of Science, Rehovot 76100, Israel. ²⁰Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. ²¹Departments of Anthropology and of Organismic and Evolutionary Biology, Harvard University, 11 Divinity Avenue, Cambridge, Massachusetts 02138, USA. ²²Department of Systems Biology, Harvard Medical School, Boston, Massachusetts 02115, USA. ²³Whitehead Institute for Biomedical Research, Cambridge, Massachusetts 02142, USA. ²⁴Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

ARTICLES

A genome-wide comparison of recent chimpanzee and human segmental duplications

Ze Cheng¹, Mario Ventura², Xinwei She¹, Philipp Khaitovich³, Tina Graves⁴, Kazutoyo Osoegawa⁵, Deanna Church⁶, Pieter DeJong⁵, Richard K. Wilson⁴, Svante Pääbo³, Mariano Rocchi² & Evan E. Eichler¹

We present a global comparison of differences in content of segmental duplication between human and chimpanzee, and determine that 33% of human duplications (>94% sequence identity) are not duplicated in chimpanzee, including some human disease-causing duplications. Combining experimental and computational approaches, we estimate a genomic duplication rate of 4–5 megabases per million years since divergence. These changes have resulted in gene expression differences between the species. In terms of numbers of base pairs affected, we determine that *de novo* duplication has contributed most significantly to differences between the species, followed by deletion of ancestral duplications. Post-speciation gene conversion accounts for less than 10% of recent segmental duplication. Chimpanzee-specific hyperexpansion (>100 copies) of particular segments of DNA have resulted in marked quantitative differences and alterations in the genome landscape between chimpanzee and human. Almost all of the most extreme differences relate to changes in chromosome structure, including the emergence of African great ape subterminal heterochromatin. Nevertheless, base per base, large segmental duplication events have had a greater impact (2.7%) in altering the genomic landscape of these two species than single-base-pair substitution (1.2%).

Recent segmental duplications have had a pivotal role in the evolution of the architecture of the human genome^{1–6}, the emergence of new genes^{7,8} and the adaptation of our species to its environment^{9–12}. They contribute to large-scale structural polymorphism^{13–17} and a host of genomic diseases¹⁸. Several gene and genomic-based analyses suggest that the human genome is particularly enriched for genes that have emerged as a result of recent duplication^{11,19}. It is unknown whether slow rates of deletion, high rates of duplication or gene conversion are largely responsible for the evolutionary maintenance of these duplicates. We sought to understand the origin and impact of this fraction of the genome by performing a detailed comparison of the human and chimpanzee genomes for regions that showed evidence of shared and lineage-specific duplication.

Chimpanzee segmental duplications

We used two independent approaches to estimate the size and extent of chimpanzee (*Pan troglodytes*) duplications. We first performed a self-comparison of the chimpanzee genome assembly using the whole-genome assembly comparison method (WGAC)²⁰. We noticed a significant (threefold) reduction of more divergent (94–95% sequence identity) chimpanzee interchromosomal pairwise alignments when compared to human (Supplementary Fig. S1). As expected, more recent duplications (>97% sequence identity) were five times as likely to be misassembled or fragmented when compared to unique chimpanzee sequence²¹. To identify these duplications using chimpanzee whole-genome shotgun (WGS) sequence data

(3.5-fold sequence coverage), we implemented a second duplication detection method¹¹ that uses the depth of coverage of random sequence read data against a reference sequence to identify duplicated sequence. We applied the whole-genome shotgun sequence detection (WSSD) strategy by mapping 23.7 million reads from chimpanzee against the human genome reference (Fig. 1). Table 1 summarizes the results of the duplication analyses for the two genomes using the WSSD approach for regions >20 kilobases (kb) in length and >94% sequence identity.

We classified DNA into one of three possible categories: duplicated only in chimpanzee, duplicated only in human or shared between chimpanzee and human (Fig. 1b–d; see Methods). On the basis of five different computational and experimental analyses, including array comparative genomic hybridization and fluorescence *in situ* hybridization (FISH) (32 out of 34 validations; Supplementary Figs S2–S6, Supplementary Tables S1–S6 and Supplementary Methods), we estimate that we have detected >90% of all segmental duplications in the chimpanzee genome that are greater than 20 kb in length. A chimpanzee segmental duplication database as well as detailed chromosomal views for patterns of human and chimpanzee duplications are available (Supplementary Fig. S7; see also <http://chimpparalogy.gs.washington.edu>) based on mapping all four duplication tracks.

Gene and duplication structure analysis

Although most (66%) of the autosomal base pairs (bp) duplicated in humans are shared between human and chimpanzee (Table 1), a

¹Howard Hughes Medical Institute, Department of Genome Sciences, University of Washington School of Medicine, 1705 NE Pacific Street, Seattle, Washington 98195, USA.

²Department of Genetics and Microbiology, University of Bari, 70126 Bari, Italy. ³Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, D-04103 Leipzig, Germany. ⁴Washington University School of Medicine, 4444 Forest Park Blvd, St Louis, Missouri 63108, USA. ⁵BACPAC Resources, Children's Hospital of Oakland Research Institute, Bruce Lyon Memorial Research Building, Oakland, California 94609, USA. ⁶National Center for Biotechnology Information, National Library of Medicine, National

Institutes of Health, Building 38A, 8600 Rockville Pike, Bethesda, Maryland 20894, USA.

surprisingly large fraction (~33% or 26.5 out of 79.8 Mb) is duplicated in human but not chimpanzee (Table 1). These human-only duplication intervals map to 515 regions with an average length of 54.6 kb; there is a particular bias for human-specific duplications noted on chromosomes 5 and 15. Significant portions of the duplication architecture that predispose humans to Williams–Beuren syndrome, juvenile nephronophthisis, spinal muscular atrophy and Prader–Willi syndrome¹⁸ appear to be single copies in the chimpanzee (Supplementary Fig. S5). Because non-allelic homologous recombination is thought to provide the molecular basis for recurrent chromosomal structural rearrangements associated with these diseases, these alterations in duplication architecture and the concomitant prevalence of these diseases may be a peculiarity of the human lineage of evolution.

From the perspective of the chimpanzee genome, we identified 11.4 Mb (202 regions) of human sequence that were duplicated in chimpanzee but not in human (approximately 224 kb of the 112 Mb of chimpanzee-specific sequence was also duplicated). If we correct for copy number in the chimpanzee genome, our analysis suggests that the two genomes show comparable levels of autosomal lineage-specific duplication (31.9 Mb in human versus 36.2 Mb in chimpanzee). In contrast, if we compare copy number estimates for shared duplications (588 regions, average length = 94.3 kb), we estimate that the chimpanzee genome has increased in size by as much as 26 Mb (1%), largely as a result of the hyperexpansion of a small number of chimpanzee segmental duplications (see below).

We determined that a chimpanzee-only and human-only duplication were 10.3 times more likely to be located in close proximity to

a shared duplication than predicted based on a random simulation model (Supplementary Table S7). These data indicate that either lineage-specific deletion or duplication is occurring in proximity to regions of shared duplication. This effect, which we have termed ‘duplication shadowing’, suggests that loci near clusters of segmental duplication may be more susceptible to duplication/deletion, probably due to an increased frequency of non-allelic homologous recombination¹⁸.

A total of 177 complete and partial genes (88 and 89 respectively) show evidence of duplication in human but not chimpanzee (for example, *SMN*, KARP1 binding protein, *N*-ethylmaleimide-sensitive factor, *CCL4L1*; Supplementary Table S8). In contrast, only 94 genes were duplicated in chimpanzee but not humans (for example, interleukin receptor-like 1, huntingtin interacting protein 1, and bone morphogenetic protein 2) (Supplementary Table S9). On the basis of U133 Affymetrix gene expression comparisons between human and chimpanzee for five tissues ($n = 30,323$ transcripts), we determined that 56% of the human-only gene duplicates showed significant differences in gene expression—83% of this gene expression difference was due to upregulation within human as opposed to chimpanzee ($P < 0.0001$). Similarly, 49% of chimpanzee duplications found in chimpanzee but not in human showed changes in gene expression within chimpanzee when compared to humans—57% was due to upregulation within chimpanzee as opposed to human ($P < 0.01$) (Supplementary Tables S10 and S11). These data indicate that a significant proportion of the lineage-specific duplications resulted in gene expression differences between the two species.

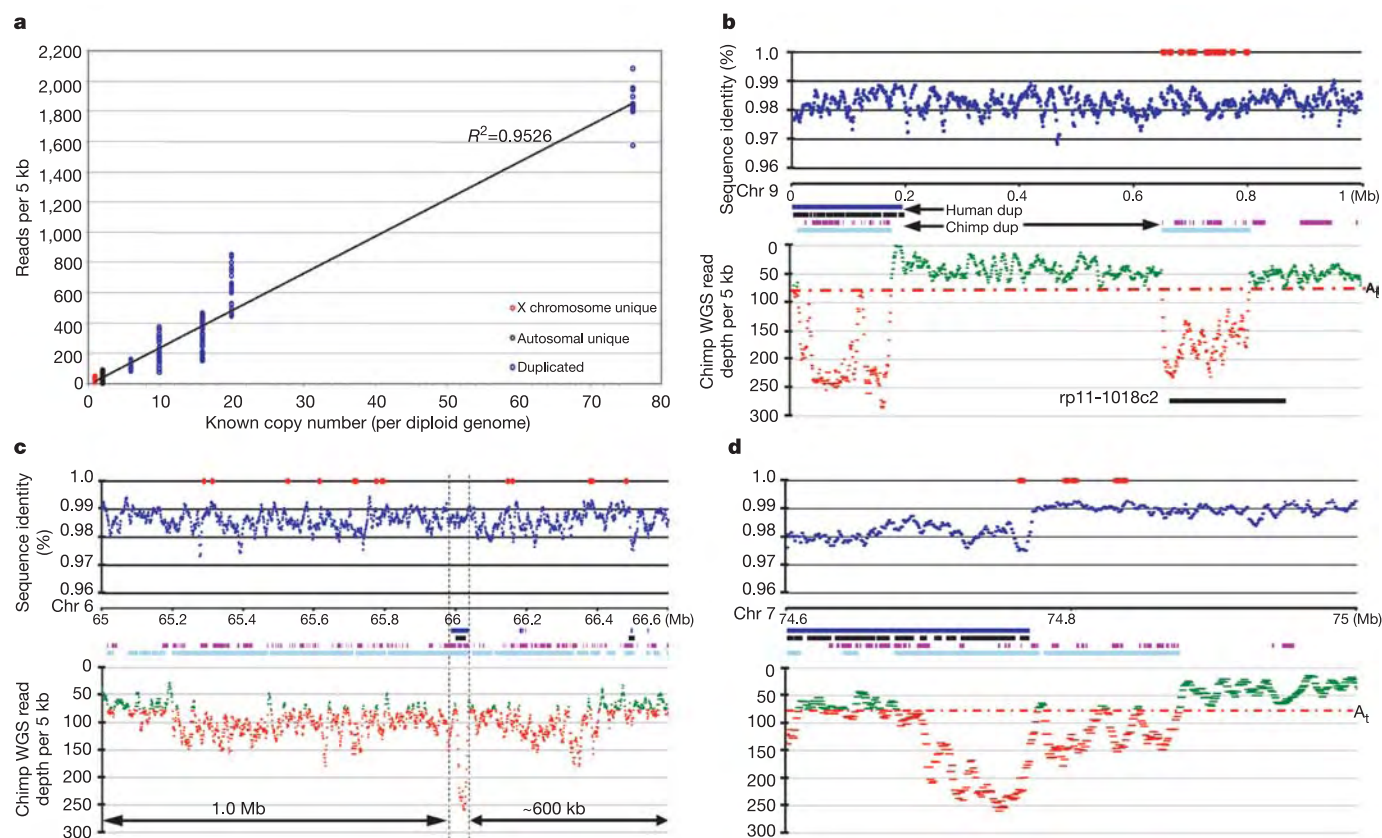


Figure 1 | Chimpanzee segmental duplication detection on human genome assembly NCBI-34 (build 34). **a**, Correlation of copy number and whole-genome shotgun sequence read coverage ($R^2 = 0.953$) is shown based on analysis of unique and duplicated chimpanzee loci of known copy number (Supplementary Table S1). **b–d**, Three examples of chimpanzee-only duplications are depicted based on comparison of the four duplication

analyses (human WGAC, dark blue; human WSSD, black; chimpanzee WGAC, purple; chimpanzee WSSD, light blue). Significant departures (3 s.d.) in the depth-of-coverage of chimpanzee reads (5-kb windows) are shown below the tracks (red). Red dots indicate the position of ‘triallelic’ variants (Supplementary Methods).

Table 1 | Chromosome distribution of chimpanzee and human segmental duplications

Chromosome	Size	Human WSSD	Chimpanzee WSSD	Shared	Human only	Chimpanzee only	Shared duplications (human copy number corrected)	Shared duplications (chimpanzee copy number corrected)	Human only duplications (copy number corrected)	Chimpanzee only duplications (copy number corrected)
Chr 1	221.56	6.88	6.13	4.18	2.71	1.96	7.09	7.94	4.17	5.62
Chr 2*‡	237.54	7.95	5.42	4.86	3.09	0.57	5.11	8.42	3.78	1.25
Chr 3	194.47	1.52	1.25	1.06	0.46	0.19	1.21	1.28	0.72	0.55
Chr 4‡	186.84	2.86	2.45	2.29	0.57	0.15	4.91	9.01	0.87	0.55
Chr 5*	177.55	4.17	2.73	2.40	1.77	0.33	3.24	3.42	2.00	0.89
Chr 6†	167.26	1.19	2.86	0.92	0.28	1.95	1.31	1.31	0.38	5.25
Chr 7*†	154.68	7.94	7.02	5.19	2.75	1.83	5.18	6.38	2.93	4.26
Chr 8	142.35	2.19	2.47	1.88	0.31	0.58	2.97	3.78	0.29	1.66
Chr 9*‡	115.62	8.44	7.32	6.61	1.83	0.71	7.25	20.76	1.81	2.16
Chr 10†	131.17	4.73	3.76	3.43	1.30	0.33	3.37	3.41	1.76	3.47
Chr 11	130.91	2.47	2.18	1.78	0.69	0.40	1.63	1.87	0.78	0.88
Chr 12	129.83	0.89	0.79	0.65	0.24	0.14	0.65	0.65	0.21	0.33
Chr 13	95.56	0.99	0.70	0.58	0.41	0.12	0.75	0.74	0.43	1.25
Chr 14	87.19	0.72	0.55	0.25	0.47	0.30	0.25	0.25	0.46	0.70
Chr 15*	81.26	6.75	3.08	2.87	3.87	0.21	2.55	2.48	4.31	0.92
Chr 16*	79.93	8.00	6.82	5.99	2.01	0.83	6.34	6.72	2.31	3.87
Chr 17*	77.68	4.48	3.36	2.99	1.49	0.37	3.23	3.21	1.50	1.15
Chr 18	74.65	1.45	1.10	1.06	0.39	0.04	1.29	1.47	0.38	0.12
Chr 19	55.79	1.43	1.10	0.97	0.46	0.13	1.36	2.60	1.07	0.96
Chr 20	59.42	0.82	0.79	0.77	0.05	0.02	1.00	0.99	0.07	0.00
Chr 21	33.92	1.52	0.97	0.94	0.58	0.03	1.16	1.10	0.86	0.09
Chr 22	34.35	2.44	1.89	1.73	0.71	0.17	2.34	2.22	0.75	0.30
Total (autosome)	2,669.55	79.84	64.75	53.39	26.45	11.36	64.18	90.02	31.86	36.22
Chr X	149.22	6.17	1.96	1.63	4.54	0.33	2.64	2.02	4.98	1.41
Chr Y	24.65	10.56	3.88	3.56	7.00	0.32	4.00	4.77	7.06	1.45
Total	2,843.42	96.57	70.59	58.58	37.99	12.01	70.83	96.82	43.89	39.08

All values are in megabases. All segmental duplications (>94% identity, >20 kb in length) detected by WSSD were compared between chimpanzee and human based on the human genome sequence reference. Intervals were compared and duplications were classified as shared, chimpanzee only and human only (Methods). Copy number correction was performed based on factoring the number of redundant (duplicated) base pairs in the human genome and the estimated copy number of duplications as determined by WGS depth of coverage. The chimpanzee donor sequence was male. A detailed view for each region is available (<http://chimparalogy.gs.washington.edu> and Supplementary Fig. S7). The average per cent lineage-specific duplication per autosome is $1.35 \pm 1.22\%$ and $1.33 \pm 1.22\%$ for human and chimpanzee, respectively.

*Chromosomes that show an excess of human-only duplications (>2.5% duplication).

†Chromosomes that show an excess of chimpanzee-only duplications (>2.5% duplication).

‡Three chromosomes (2, 4 and 9) account for 16 Mb of the increase (25.8 Mb in total) in shared autosomal duplication content in chimpanzee.

Rate estimates

Three possible scenarios have been put forward to explain the 'excess' of segmental duplications within the human–ape lineage when compared to other genomes^{22,23}: frequent *de novo* duplication, a slow culling of duplications by deletion, and/or extensive gene conversion of ancient duplications^{11,24,25}. Cross-species comparison of the chimpanzee-only duplications among humans and the great apes revealed that the majority (11 out of 17) of the duplications were restricted to the chimpanzee (multiple hybridization signals were not observed in human, gorilla or orang-utan; Supplementary Table S12). These probably emerged as a consequence of *de novo* segmental duplication after speciation. Six out of seventeen of the chimpanzee-only duplications, however, were also duplicated in the gorilla (and in one case orang-utan). We propose that these apparent duplications arose before the divergence of humans and great apes and have been subsequently deleted within the human lineage, although a small fraction of these (~30%) are expected to be due to lineage-specific sorting in the ancestral chimpanzee–gorilla population²⁶.

To address further this question and the potential for gene conversion²⁵, we compared the divergence patterns of shared chimpanzee and human duplications and human-only duplications. For the former case, we limited our analysis to those where only two copies of the duplications existed (binary duplicate patterns) (Supplementary Fig. S8). Using an estimate for chimpanzee–human sequence divergence (0.0131 ± 0.0045 nucleotide substitutions per site, Supplementary Table S13), we classified duplications as occurring before, after or near the time of speciation (Fig. 2; see also Supplementary Table S14). If one examines shared duplications, we note a very small fraction, ~8% by base pairs (3% by count), with a sequence identity consistent with post-speciation gene conversion events. For human-only duplications, 67.0% of the 'new' duplication base pairs show divergence consistent with a *de novo* duplication,

whereas the remainder are more divergent, suggesting deletion of a more ancient duplication. Similar results were obtained if chimpanzee-only duplications were considered, although the number of alignments is larger due to the fragmented nature of the chimpanzee genome assembly. These findings closely parallel the results obtained by FISH and suggest that, at the base pair level, *de novo* duplication

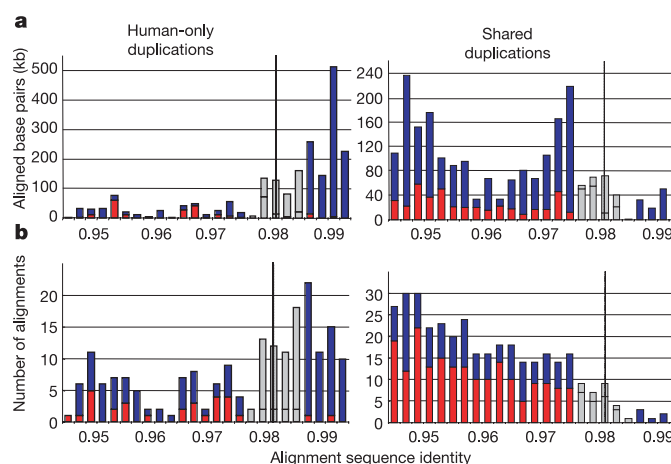


Figure 2 | Sequence identity spectra of human only versus shared duplications. **a, b,** The sequence identity (0.2% increments) of human only and shared duplication alignments is shown as a function of the total number of base pairs (**a**) and by count (**b**). Only single pairwise alignments were considered for shared duplications (Supplementary Fig. S8). Shared duplications were supported by human WGAC and either chimpanzee WSSD or chimpanzee WGAC. Interchromosomal (red), intrachromosomal (blue) and duplication alignments that lie within 1 s.d. of the chimpanzee–human sequence divergence (grey) are shown (Supplementary Table S13).

followed by deletion have contributed most significantly to the abundance of large, highly identical duplications within the human genome and not gene conversion of older duplications.

Hyperexpansion of chimpanzee segmental duplications

We examined all duplications for copy number differences between human and chimpanzee by estimating their representation in both genomes using the whole-genome shotgun sequencing detection method¹¹. Regression analysis between duplications of known copy number and the depth of random sequence show excellent correspondence in both chimpanzee and human genome sequence libraries ($r^2 = 0.953$ and $r^2 = 0.96$, respectively) (Fig. 1a). For every shared duplication interval, we computed the differential copy number (δ) for each 5-kb window of human sequence (July 2003 Assembly) after correction for common repeat sequences (Methods). We limited our analysis to duplication intervals where ten or more consecutive windows (~ 14 kb in length) showed a copy number difference in either species greater than five ($\delta > 5$).

A total of 296 regions (7.2 Mb) were identified where the human genome showed significant increases in copy number when compared to chimpanzee (Supplementary Table S15). Thirty-three per cent of the human increase (98 out of 296 intervals) mapped within 5 Mb of the centromere, corresponding to 21 out of 29 pericentromeric duplication regions in the human genome. This was significantly different ($P = 0.0002$, Fisher's exact test) compared with the chimpanzee genome, which showed relatively little increase in pericentromeric duplication intervals (13 out of 92). Array comparative genomic hybridization (CGH) between human and chimpanzee genomes (Supplementary Fig. S6) confirms these results and suggests either a genome-wide global expansion of pericentromeric duplications in the human lineage or deletion of such duplication in the chimpanzee lineage.

In contrast, we identified only 92 regions (45 clusters) where the chimpanzee genome showed a significant increase in copy number when compared to human. Although these regions are fewer in number, they correspond to a more marked increase in the amount (22.6 Mb) of shared duplicated sequence that has occurred in the chimpanzee lineage. More than 70% (16.0 out of 22.6 Mb) of the chimpanzee increase mapped to two clusters on chromosomes 2, 4 and 9. One segmental duplication in particular was identified that showed an extraordinary increase in chimpanzee when compared to

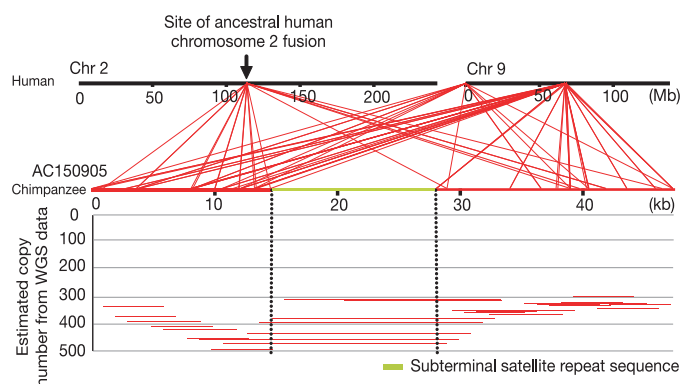


Figure 3 | Sequence structure of chimpanzee subterminal duplication. A schematic diagram depicting the organization of chimpanzee BAC 100G12 (AC150905) is shown. Segmental duplications (red) flank a 32-bp subterminal satellite repeat sequence associated with subterminal portions of great ape chromosomes. A large excess of chimpanzee reads map (on average 20,000 reads with 99.2% sequence identity) to each 5 kb of 'unique' sequence within the duplications, indicative of 300–500 copies of the segmental duplication. By comparison, the human genome assembly shows only four to five copies of this sequence, mapping to 9p24, 9q21 and 2q12.

human. Our analysis indicated that this locus (~ 40 kb in size) mapped to four regions in human but was represented ~ 400 times within the chimpanzee genome (Fig. 3). The copies in human showed high sequence identity (99.2%) and mapped to human chromosome 9p24, 9q21 and near the ancestral centromere on 2q12.

Comparative FISH analysis (Fig. 4; see also Supplementary Fig. S9) revealed that the *Pan* hyperexpansion occurred in the common ancestor of bonobo (*Pan paniscus*) and chimpanzee (*Pan troglodytes*) (2–5 million years ago) and was targeted, with the exception of interstitial regions on chromosomes VII and XIII (phylogenetic group designation), exclusively to the subterminal portions of chimpanzee chromosomes. Sequence analysis of one chimpanzee locus (Fig. 3; AC150905) revealed the presence of a single copy of the 36-kb segmental duplication and a 14.5-kb cluster of tandem repetitive repeats (32-bp repeat unit clustered into larger 400–800-bp structures). Sequence similarity searches showed significant sequence identity with previously described subterminal satellite repeats (pCHT7 and pCHT13)²⁷.

We propose that most of the asymmetrical increase of duplicated DNA in the chimpanzee lineage has emerged as a mechanistic consequence of changes in chromosome structure and not selection. The subterminal caps are an idiosyncratic structural aspect of African great ape chromosomes²⁸, which are generally regarded as heterochromatic. Similar to human pericentromeric DNA, the regions have served as sinks for duplicative transposition and expansion of particular euchromatic segments. This process has led to an overall increase in chimpanzee genome size of at least 16 Mb since human and chimpanzee separated. It is interesting that the same region that represents the site of chromosome 2 fusion²⁹ in the human lineage has undergone a segmental duplication hyperexpansion within the subterminal region of chimpanzee chromosomes. This may suggest an inherent instability of this segment of DNA, further extending the association of segmental duplication and chromosomal rearrangement without a direct cause and effect relationship¹.

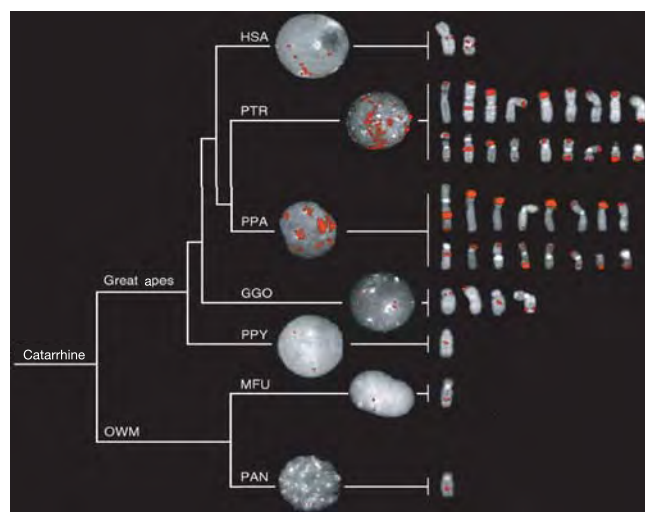


Figure 4 | A chimpanzee hyperexpansion of a shared segmental duplication. A human fosmid DNA clone (WIBR2-1785A6) corresponding to the duplicated region was hybridized against a series of primate chromosomes at metaphase, including human, common chimpanzee (*P. troglodytes*), bonobo (*P. paniscus*), gorilla, orang-utan, macaque and baboon. Hundreds of copies map to the subterminal portions of only chimpanzee and bonobo chromosomes, indicating a lineage-specific duplication expansion 2–6 million years ago. Interstitial chromosome signals are also noted on chromosomes VII and XIII and correspond to cross-hybridization with subterminal satellite repeat sequence. GGO, *Gorilla gorilla*; HSA, *Homo sapiens*; MFU, *Macaca fuscata*; OWM, Old World monkey; PAN, *Papio anubis*; PPA, *Pan paniscus*; PPY, *Pongo pygmaeus*; PTR, *Pan troglodytes*.

Discussion

Our analysis has revealed some important properties regarding the emergence and maintenance of segmental duplications within the human–ape lineage. First, we have determined that a significant fraction of the genome (1.5% or 46 Mb) is specifically duplicated in one lineage but not the other (Methods). Second, both FISH and sequence divergence data indicate that ~60% of these apparent lineage-specific differences are the result of *de novo* duplications, whereas most of the remainder is the result of deletion. In contrast to recent studies of the Y chromosome²⁵, the impact of gene conversion appears minimal for binary duplicates (<10%). Finally, in addition to qualitative differences in duplication content, we have identified significant (>5) copy number differences among shared human and chimpanzee duplications. These differences have contributed to a net gain of ~26 Mb of segmental duplication within the chimpanzee lineage (Table 1). In total, we conservatively estimate that 70 Mb (2.7%) of euchromatic sequence have been differentially duplicated between the chimpanzee and human, with 4.4 Mb of new genetic material being added on average per million years. Owing to limitations in our genomic duplication detection strategy (>20 kb), ours is almost certainly an underestimate (Methods). Nevertheless, when compared to single-base-pair differences, which account for 1.2% genetic difference, base per base, large segmental duplication events have had a greater impact (2.7%) in altering the genomic landscape of these two species.

METHODS

Duplication analyses. To detect chimpanzee duplications (>1 kb and >90% sequence identity), we performed a WGAC²⁰ on the Arachne November 2003 chimpanzee genome assembly²¹. We detected a total of 51,573 pairwise alignments corresponding to 136.7 Mb (35,453 non-redundant fragments) of ‘duplicated’ material. Forty per cent (54.3 Mb) of these fragments localized to unmapped portions of the chimpanzee genome (random assignment). As a second measure of chimpanzee duplication, independent from the genome assembly comparison, we modelled the depth of coverage of chimpanzee WSSD (23.7 million sequence reads)¹¹ against the human genome reference. The number of reads within 5-kb windows correlated strongly with copy number of duplication ($r^2 = 0.953$) (Fig. 1a). We set our thresholds of duplication detection at 75 reads per 5 kb for autosomes and 44 reads per 5 kb for the sex chromosomes (3 s.d. beyond the mean coverage based on our analysis of unique sequence). We defined a WSSD duplication interval as any region where six out of seven continuous windows showed read depth in excess of autosomal and sex chromosome thresholds. We focused on WSSD regions >20 kb in length (70.6 Mb in total) due to estimated false positive (<1.4%) and negative rates (<6.5%) at this length cut-off (Supplementary Figs S2 and S3). A corresponding chimpanzee segmental duplication database and UCSC genome browser track (<http://chimparrayology.gs.washington.edu>) were developed.

WGAC and WSSD duplication intervals were compared between human and chimpanzee by mapping all four tracks onto the human genome reference. We initially categorized DNA as duplicated in human or chimpanzee based on a comparison of these four duplication analyses (chimpanzee WGAC, human WGAC, chimpanzee WSSD and human WSSD) (Fig. 1b–d). Chimpanzee duplication intervals were defined on the human reference genome as the longest interval of contiguous duplication that seeded within at least 20 kb of chimpanzee WSSD (see below). We similarly limited our analysis of human duplications to regions of >94% sequence identity and >20 kb in length. Regions were classified into one of three possible categories: duplicated only in chimpanzee, duplicated only in human or shared between chimpanzee and human.

Validation. Five separate analyses were performed to validate our database of chimpanzee segmental duplication and to provide estimates for false positive and negative detection rates³⁰ (Supplementary Methods, Supplementary Fig S2–S6 and Supplementary Tables S1–S6).

Expression analysis. Gene expression differences between human and chimpanzee were assessed for five tissues (heart, brain, liver, testis and kidney) using Affymetrix HG U133plus2 arrays (see Supplementary Methods) as described³¹. All primary expression data are publicly available at ArrayExpress (<http://www.ebi.ac.uk/arrayexpress/>), accession number E-AFMX-11.

Rate estimate. We estimated the amount of human-only (26.5 Mb) and chimpanzee-only (11.4 Mb) duplication and adjusted for copy number based on the WGS depth-of-coverage estimate (WSSD) of each corresponding region

(31.9 Mb and 36.2 Mb, respectively) (Table 1). The amount of new duplicated material in chimpanzee was then simply $36.2 - 11.4$ Mb (24.8 Mb), whereas the amount of new human autosomal material was corrected for the copy number of the reference human genome assembly ($31.9 - (26.5/2.6) = 21.7$ Mb). We estimate that there has been a minimum of 46.5 Mb of lineage-specific segmental duplication since separation of chimpanzee and human. We determined that there has been an increase of 7.2 Mb and 22.6 Mb of shared (chimpanzee and human) duplication in the human and chimpanzee lineages, respectively, for regions where the genomic copy number increased by five or more (Supplementary Table S15). Sixteen megabases of the chimpanzee increase is due to a lineage-specific expansion that occurred before the separation of *P. troglodytes* from *P. paniscus* (2 million years ago), but after the separation of human and chimpanzee (6 million years ago); 13.8 Mb (7.2 Mb in human and 6.6 Mb in chimpanzee) is the result of marked changes in copy number of a subset of segmental duplications between the two species. If we estimate that 60% of these bases have emerged by duplication, as opposed to deletion and gene conversion (Supplementary Table S14), we calculate that $0.6(13.8 + 46.5) + 16$ Mb = 52.2 Mb have arisen as a result of *de novo* duplication since divergence of the two species. This corresponds to 4.4 Mb of duplication per million years or an effective fixation rate of 3.4 Mb of segmental duplication per million years (assuming chimpanzee–human separation at 6 million years ago and a polymorphism frequency of 0.2). This rate is a lower bound estimate, because sex chromosome duplications as well as autosomal duplications <20 kb in size were not considered owing to reduced power of detection in the chimpanzee lineage. If we extrapolate based on the analysis of duplications in human (Supplementary Table S2), we can calculate an upper bound of *de novo* duplication of 5.5 Mb per million years.

Received 31 March; accepted 30 June 2005.

- Bailey, J. A., Baertsch, R., Kent, W. J., Haussler, D. & Eichler, E. E. Hotspots of mammalian chromosomal evolution. *Genome Biol.* **5**, R23 (2004).
- She, X. *et al.* The structure and evolution of centromeric transition regions within the human genome. *Nature* **430**, 857–864 (2004).
- Armengol, L., Pujana, M. A., Cheung, J., Scherer, S. W. & Estivill, X. Enrichment of segmental duplications in regions of breaks of synteny between the human and mouse genomes suggest their involvement in evolutionary rearrangements. *Hum. Mol. Genet.* **12**, 2201–2208 (2003).
- Trask, B. *et al.* Members of the olfactory receptor gene family are contained in large blocks of DNA duplicated polymorphically near the ends of human chromosomes. *Hum. Mol. Genet.* **7**, 13–26 (1998).
- Eichler, E. E. *et al.* Duplication of a gene-rich cluster between 16p11.1 and Xq28: a novel pericentromeric-directed mechanism for paralogous genome evolution. *Hum. Mol. Genet.* **5**, 899–912 (1996).
- Ventura, M. *et al.* Neocentromeres in 15q24–26 map to duplicons which flanked an ancestral centromere in 15q25. *Genome Res.* **13**, 2059–2068 (2003).
- Johnson, M. E. *et al.* Positive selection of a gene family during the emergence of humans and African apes. *Nature* **413**, 514–519 (2001).
- Courseaux, A. & Nahon, J. L. Birth of two chimeric genes in the Hominidae lineage. *Science* **291**, 1293–1297 (2001).
- Stefansson, H. *et al.* A common inversion under selection in Europeans. *Nature Genet.* **37**, 129–137 (2005).
- Gonzalez, E. *et al.* The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science* **307**, 1434–1440 (2005).
- Bailey, J. A. *et al.* Recent segmental duplications in the human genome. *Science* **297**, 1003–1007 (2002).
- Khaitovich, P. *et al.* Regional patterns of gene expression in human and chimpanzee brains. *Genome Res.* **14**, 1462–1473 (2004).
- Sebat, J. *et al.* Large-scale copy number polymorphism in the human genome. *Science* **305**, 525–528 (2004).
- lafrate, A. J. *et al.* Detection of large-scale variation in the human genome. *Nature Genet.* **36**, 949–951 (2004).
- Tuzun, E. *et al.* Fine-scale structural variation of the human genome. *Nature Genet.* **37**, 727–732 (2005).
- Sharp, A. J. *et al.* Segmental duplications and copy number variation in the human genome. *Am. J. Hum. Genet.* **77**, 78–88 (2005).
- Fredman, D. *et al.* Complex SNP-related sequence variation in segmental genome duplications. *Nature Genet.* **36**, 861–866 (2004).
- Stankiewicz, P. & Lupski, J. R. Genomic architecture, rearrangements and genomic disorders. *Trends Genet.* **18**, 74–82 (2002).
- Zhang, L., Lu, H. H., Chung, W. Y., Yang, J. & Li, W. H. Patterns of segmental duplication in the human genome. *Mol. Biol. Evol.* **22**, 135–141 (2005).
- Bailey, J. A., Yavor, A. M., Massa, H. F., Trask, B. J. & Eichler, E. E. Segmental duplications: organization and impact within the current human genome project assembly. *Genome Res.* **11**, 1005–1017 (2001).
- The Chimpanzee Sequencing and Analysis Consortium. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* doi:10.1038/nature04072 (this issue).

22. Tuzun, E., Bailey, J. A. & Eichler, E. E. Recent segmental duplications in the working draft assembly of the brown Norway rat. *Genome Res.* **14**, 493–506 (2004).
23. Bailey, J. A., Church, D. M., Ventura, M., Rocchi, M. & Eichler, E. E. Analysis of segmental duplications and genome assembly in the mouse. *Genome Res.* **14**, 789–801 (2004).
24. International Human Genome Sequencing Consortium, Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931–945 (2004).
25. Rozen, S. *et al.* Abundant gene conversion between arms of massive palindromes in human and ape Y chromosomes. *Nature* **423**, 873–876 (2003).
26. Chen, F. C. & Li, W. H. Genomic divergences between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees. *Am. J. Hum. Genet.* **68**, 444–456 (2001).
27. Royle, N. J., Baird, D. M. & Jeffreys, A. J. A subterminal satellite located adjacent to telomeres in chimpanzees is absent from the human genome. *Nature Genet.* **6**, 52–56 (1994).
28. Yunis, J. J. & Prakash, O. The origin of man: a chromosomal pictorial legacy. *Science* **215**, 1525–1530 (1982).
29. Fan, Y., Linardopoulou, E., Friedman, C., Williams, E. & Trask, B. J. Genomic structure and evolution of the ancestral chromosome fusion site in 2q13–2q14.1 and paralogous regions on other human chromosomes. *Genome Res.* **12**, 1651–1662 (2002).
30. Fortna, A. *et al.* Lineage-specific gene duplication and loss in human and great ape evolution. *PLoS Biol.* **2**, E207 (2004).
31. Khaitovich, P. *et al.* Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. *Science* (in the press).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Lachmann, I. Hellman and G. Vessere for technical assistance; the Chimpanzee Sequencing and Analysis Consortium for access to the chimpanzee sequence data before publication; A. Force for discussions; and J. Pecotte, S. Warren and J. Rogers for providing some of the primate material used in this study. This work was supported by grants from the National Human Genome Research Institute, the National Institute of General Medical Sciences, Centro di Eccellenza Geni in campo Biosanitario e Agroalimentare, Ministero Italiano della Università e della Ricerca, the European Commission and the Bundesministerium für Bildung und Forschung.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to E.E.E. (eee@gs.washington.edu).

ARTICLES

Human subtelomeres are hot spots of interchromosomal recombination and segmental duplication

Elena V. Linardopoulou^{1,2}, Eleanor M. Williams¹, Yuxin Fan^{1†}, Cynthia Friedman¹, Janet M. Young¹ & Barbara J. Trask^{1,2,3}

Human subtelomeres are polymorphic patchworks of interchromosomal segmental duplications at the ends of chromosomes. Here we provide evidence that these patchworks arose recently through repeated translocations between chromosome ends. We assess the relative contribution of the principal mechanisms of ectopic DNA repair to the formation of subtelomeric duplications and find that non-homologous end-joining predominates. Once subtelomeric duplications arise, they are prone to homology-based sequence transfers as shown by the incongruent phylogenetic relationships of neighbouring sections. Interchromosomal recombination of subtelomeres is a potent force for recent change. Cytogenetic and sequence analyses reveal that pieces of the subtelomeric patchwork have changed location and copy number with unprecedented frequency during primate evolution. Half of the known subtelomeric sequence has formed recently, through human-specific sequence transfers and duplications. Subtelomeric dynamics result in a gene duplication rate significantly higher than the genome average and could have both advantageous and pathological consequences in human biology. More generally, our analyses suggest an evolutionary cycle between segmental polymorphisms and genome rearrangements.

The human genome contains an abundance of large DNA segments that have duplicated over the last 40 million years^{1,2}. These segmental duplications represent $\geq 5\%$ of the genome² and are frequently found near centromeres and telomeres³. Segmental duplications are emerging as important factors in chromosomal rearrangements leading to disease⁴ and rapid gene innovation², but the mechanisms by which they form are not well understood. Here we focus on the unusually dense concentrations of interchromosomal segmental duplications comprising human subtelomeres, which form the transition zones between chromosome-specific sequence and the arrays of telomeric repeats capping each chromosomal end. Previous cytogenetic studies have shown that human subtelomeres are strikingly polymorphic in content—large segments can be present in or absent from normal alleles⁵—and that the copy number of subtelomeric segments can vary among higher primates^{6–9}. This natural plasticity, combined with documented expression of several human subtelomeric genes^{10,11}, suggests that the evolutionary dynamics of subtelomeric regions could contribute to normal phenotypic variation within and between primate species, as is observed in other organisms (reviewed in ref. 5). However, subtle rearrangements of DNA near the ends of chromosomes are observed in association with human disorders, including mental retardation¹². Although full sequence coverage has not yet been achieved for all chromosome ends, let alone for multiple alleles of each end, much can be learned from available sequence about subtelomere organization, evolution, variation and function, as well as more generally about the origin and consequences of segmental duplications.

Complex interrelated structures

Our ‘paralogy map’ of subtelomeric segmental duplications (Fig. 1 and Supplementary Table S1) uses all finished sequences of genomic clones submitted to GenBank before April 2003. The map comprises ~ 2.6 Mb of sequence present in two or more of 33 human subtelomeres (including three allelic pairs). The seven completely sequenced subtelomeres in the set are bounded distally by 0.5–2.4 kb of various tandemly repeated units¹³ called telomere-associated repeats (TAR1) and a short sample of the native telomeric arrays¹⁴. Numerous degenerate telomere-like repeats and TAR1 elements are also situated at varying distances from telomeres^{15,16} (Fig. 1). Notably, these repeats are almost always oriented 5′–3′ towards the telomere.

The paralogy map reveals the complex patchwork of sequence blocks shared by human subtelomeres. Different subtelomeres can show >100 kb continuous similarity, but a segment shared by a given chromosome set extends only 13 kb on average before being displaced on at least one subtelomere by a segment with a different chromosomal distribution. In the 33 subtelomeric contigs analysed, we identify 41 homology blocks larger than 3 kb (Fig. 1) (blocks 42–44 are special cases and not counted, see Supplementary Table S2). These blocks occur in 2–18 copies (average of 5), with 88–99.9% identity (Supplementary Table S2). Almost all instances of these blocks are in the same orientation and relative order (Fig. 1). Polymerase chain reaction (PCR) analyses of monochromosomal hybrid cell lines confirm block boundaries defined by sequence

¹Division of Human Biology, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North C3-168, Seattle, Washington 98109, USA. ²Department of Bioengineering, University of Washington, Box 357962, Seattle, Washington 98195-7962, USA. ³Department of Genome Sciences, University of Washington, Box 357730, Seattle, Washington 98195-7730, USA. †Present address: Departments of Laboratory Medicine and Medicine (Division of Medical Genetics), University of Washington, Seattle, Washington 98195, USA.

alignments and identify at least one additional chromosomal copy for 17 out of 29 blocks evaluated (Supplementary Fig. S1 and Table S3).

Subtelomeres contain members of 25 small families of genes (Fig. 1). There is one gene per 30 kb on average. Eighteen families contain at least one subtelomeric member that encodes a potentially functional protein (Supplementary Table S4). Thus, gain, loss or alteration of subtelomeric genes has a potential phenotypic effect. Subtelomeric genes have highly varied functions and include odorant and cytokine receptors, tubulins, transcription factors and genes of unknown function.

Sequences in the paralogy map and duplicates thereof detected in later assemblies and/or by PCR (an added total of 0.97 Mb) account for $\geq 83\%$ of the estimated subtelomeric terrain in a typical genome¹⁶. Approximately 90% of the 490 kb of finished sequence added to nine ends in the latest genome assembly (Build 35) is $>90\%$

identical to sequence already represented in our data set; only 26 kb is novel. Thus, our data set represents a reasonably comprehensive sample from which mechanistic information can be derived.

Mechanisms of sequence transfer

To investigate the mechanisms resulting in subtelomeric segmental duplications, we considered two phases in their evolutionary history. The first consists of duplication to a new chromosome, creating a new structural boundary, and the second involves possible interactions between existing duplicates. We analysed the patterns and breakpoints of homology in sequenced subtelomeres to infer the mechanism of interchromosomal sequence transfer that would result in the first step. Two primary models were considered that might give rise to subtelomeric segmental duplications, namely chromosome translocations and DNA transposition.

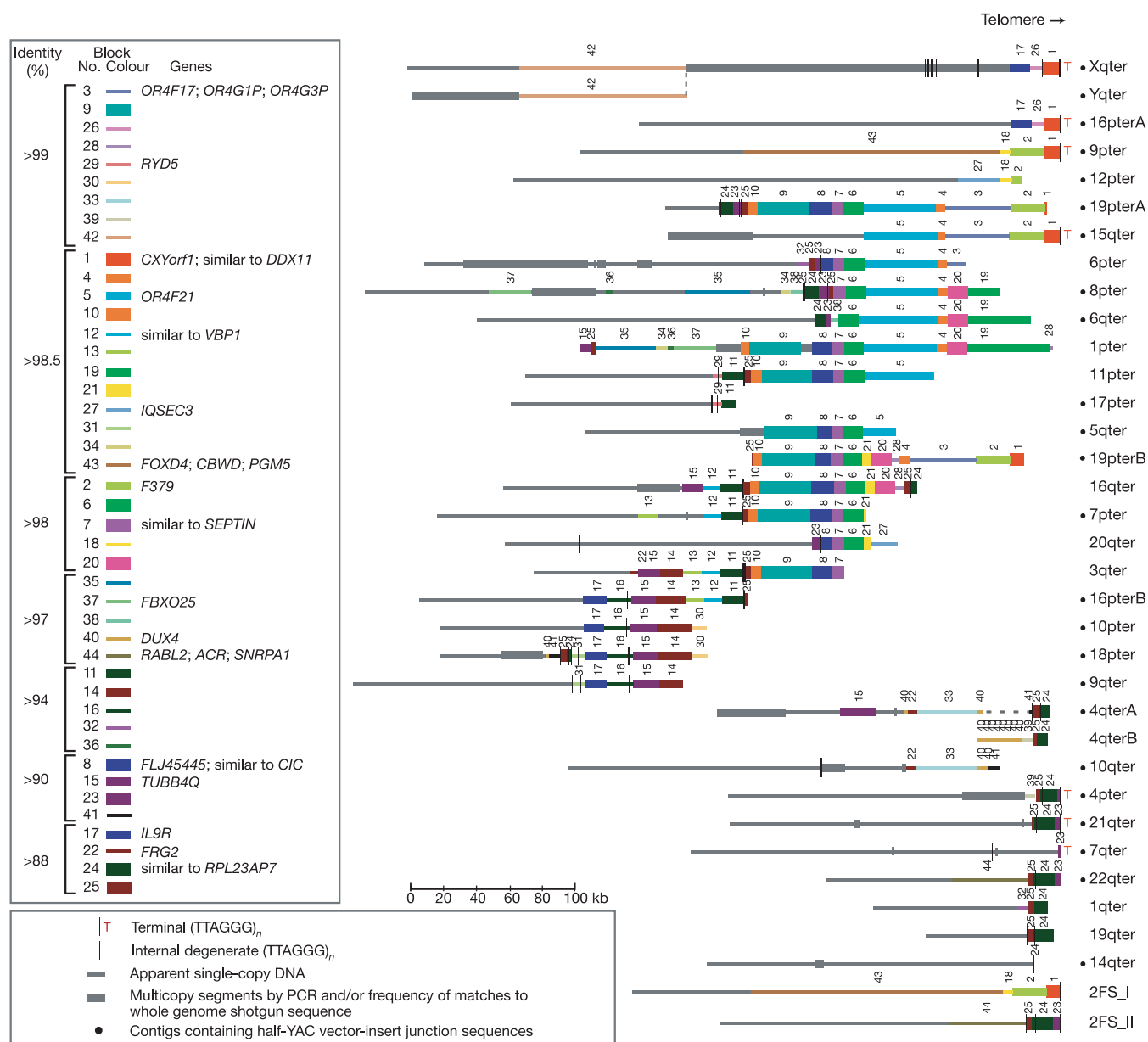


Figure 1 | Subtelomeric paralogy map. Subtelomeric contigs (Supplementary Table S1 lists constituent accession numbers and localization methods) are aligned at telomeres or to maximize alignments of paralogous blocks. Copies of a given block have the same colour, line width and number. Only blocks 15 and 40 on 4q, 22 on 3q, 34–37 on 1p, and 38 on 6q are in inverted orientation relative to other corresponding block copies.

2qFS_I and _II represent ancestral telomeres fused head-to-head at 2q13-14; other internal paralogies are not displayed or analysed here. A and B indicate allelic variants. Yq/Xq pseudoautosomal homology extends distal of dotted line. See Supplementary Table S4 for details about the subtelomeric gene copies.

Several observations argue for the translocation model (Fig. 2) and against transposition. First, subtelomeric blocks do not have characteristic features associated with known transposons or their insertion sites¹⁷. Proposed targets for insertions of segmental duplications by a more general transpositional model¹⁸ are also not found at subtelomeric homology breakpoints. Second, the preserved centromere–telomere orientation and order of most duplicated blocks and degenerate telomeric repeats, and the embedded patterns of shared blocks (Fig. 1), argue against a transpositional model.

Instead, the block patterns are consistent with patchwork formation by numerous translocations involving the tips of chromosomes and subsequent transmission of unbalanced chromosomal complements to offspring (Fig. 2). In this model, each translocation event has the potential to create a new homology boundary and define a new block. Figure 3 illustrates how two translocations led to the duplication of a subtelomeric segment (block 4 plus 5) and its juxtaposition between different neighbours on chromosomes 15q and 8p. The sequence of events can be inferred from the state of interspersed repeat elements at homology breakpoints. Chromosomes 15q and 16q represent ancestral states, and the intermediate state of 6p reveals temporal separation of the two translocations leading to the block configuration on 8p.

Translocations can result from aberrant repair by either non-homologous end-joining (NHEJ) or homologous recombination; both are major mechanisms of double-strand break (DSB) repair in mammalian cells^{19,20}. To deduce the relative contribution of NHEJ and non-allelic homologous recombination (NAHR) to subtelomeric block juxtapositions, we examined all homology breakpoints at single-nucleotide resolution. The presence of repetitive elements of the same class (or paralogous genes) at the homology boundary in

both aligned junction sequences, often with a transition from high to lower sequence identity within the repeat, is strongly suggestive of homology-based repair (for example, see Supplementary Fig. S2, where the original state can be recognized by characteristic direct repeats flanking the *Alu* element). In contrast, the absence of aligned repeats or the presence of a truncated repetitive element (or gene) at the homology boundary in one sequence is indicative of non-homologous end-joining (as at breakpoints A and B, Fig. 3).

We identified a complete, non-redundant set of 56 junction-sequence alignments, each representing a unique translocation event, in the sequenced subtelomeres. We deduced the repair mechanism in 53 of these cases (Fig. 4, Supplementary Fig. S3 and Table S5). The vast majority seem to result from NHEJ (49/53, 92%) (Fig. 4b). We infer repeat-mediated NAHR for only four (8%) of the events (Fig. 4a), three of which involved *Alu* repeats. In the 15 cases of NHEJ for which structures representing both original partners and one translocation derivative were available, we found ≤ 5 bp of homology between the original sequences at the junction site (for example, Fig. 3b). Small insertions found at eight junction sites are consistent with NHEJ-mediated translocations; eight cases of apparent large deletions could have formed either by translocation or intrachromosomal deletion (Supplementary Table S5 and Fig. S4).

Although duplication borders of segmental duplications were found in genome-wide analyses to be enriched in recently active *Alu* repeats^{21,22}, interspersed repeats are not enriched at the DSBs leading to subtelomeric segmental duplications. Of a total of 102 independent DSBs (Fig. 4), 45% occurred within a repetitive element (10.8% in *Alu* elements), close to the frequency expected from subtelomeric repeat content (Supplementary Table S6). We do, however, find degenerate telomeric repeats at 4% of these DSBs, whereas they occupy 0.5% of subtelomeric sequence (Supplementary Tables S5 and S6). Subtelomeres are notably enriched in degenerate telomeric repeats relative to adjacent single-copy sequence or other genomic regions (~ 10 - and ~ 100 -fold, respectively) (Supplementary Table S6). These repeats could have been appended during DSB repair, the postulated genesis of other interstitial telomere-like repeats²³. Breakpoint 22 in Supplementary Table S5 is a clear example of such a process. Although we cannot rule out a functional role for these repeats, they are probably scars of many past DSB repairs.

Generation of diverse structures by multiple translocations between chromosome ends (Fig. 2a–c) is just one aspect of subtelomeric dynamics. Once duplicates exist on different chromosomes, they are subject to homology-based reciprocal or non-reciprocal

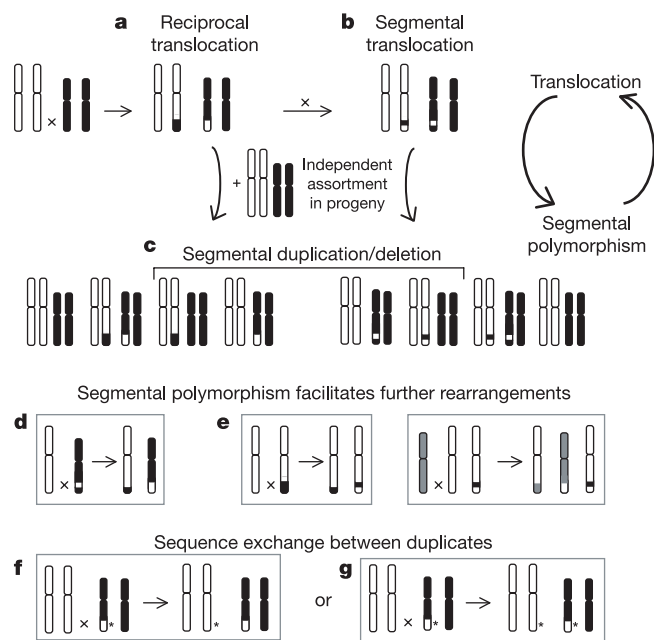


Figure 2 | A translocation-based model of segmental duplication and polymorphism. **a**, A terminal duplication/deletion can arise if a translocation product and an intact homologue are passed from parent to offspring, creating a segmental polymorphism (**c**). **b**, A segmental duplication/deletion can arise if a second interchromosomal exchange occurs between the translocated chromosomes. **d**, **e**, Segmental polymorphism can facilitate further rearrangements by promoting translocations through interchromosomal homologies (**d**) or causing translocation or other rearrangement owing to the absence of homology (**e**). **f**, **g**, Both reciprocal and non-reciprocal homology-based sequence transfers are possible between duplicates generated by any of the above steps. Asterisk indicates sequence variant.

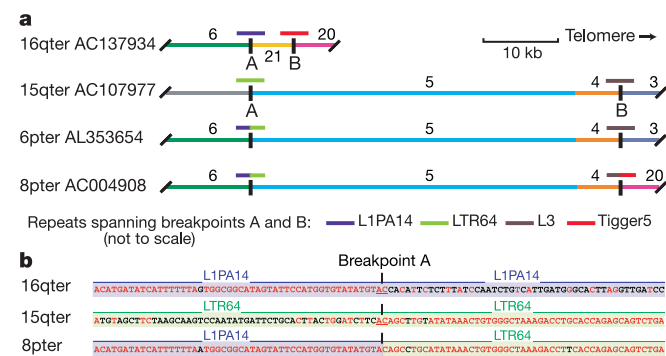


Figure 3 | Layers of interchromosomal translocations form subtelomeric blocks. **a**, Paralogous blocks have shared colour and number; short coloured lines above indicate different repetitive elements at homology breakpoints A and B, which define two translocations. An intact copy of each repeat is preserved in 16q and 15q sequences spanning the homology breakpoints with 6p and 8p, which contain truncated repeats fused by NHEJ. **b**, Only two identical nucleotides (underlined) are found at the point where the original two sequences were joined at breakpoint A to form a hybrid. Aligned bases are red.

sequence transfers (Fig. 2f, g). These events do not generate new block boundaries, but can supplant mutations accrued on one chromosome with those from another copy and spread structures formed by NHEJ to new locations (Fig. 2d). We reasoned that if duplication and subsequent homology-based sequence transfer are separated by sufficient time, the latter could be observed as a significant shift in sequence identity within regions of similarity.

To test for such events, we evaluated fluctuations in sequence identity along a 60-kb region, parts of which are shared with 88–99.5% identity by seven subtelomeres (Fig. 5a). Four computational approaches indicate that homology-based sequence transfers occurred many times between these paralogues. (1) The best-matching pairs (that is, partners in the most recent transfer events) change ≥ 5 times along the sequences (Fig. 5b and Supplementary Fig. S5). (2) The phylogenetic relationships of neighbouring sections are strikingly incongruent (Fig. 5c). (3) The per cent identity between any two subtelomeres shifts significantly multiple times across their alignment (Figs 5d, e). High similarity is unlikely to result from local selective pressure, because the most similar portions of different sequence pairs do not coincide. (4) Strong statistical support for multiple sequence transfers, ranging from several hundred to several thousand base pairs, is obtained using GeneConv²⁴ (Fig. 5f and Supplementary Table S7). Thus, subtelomeric blocks on different chromosomes do not evolve independently; instead, continued interchromosomal interactions obfuscate their duplication history. Transfers are also likely to be prevalent among the many subtelomeric blocks that are $>98\%$ identical, but only more subtle haplotype analyses might detect these events²⁵.

Dynamics of primate subtelomeres

Recent changes in subtelomeric composition can be detected using fluorescence *in situ* hybridization (FISH) to determine the copy number and location of sequences in chromosomes of different primate species. Previous descriptions of subtelomeric dynamics

Mechanism of DSB repair	Theoretical translocating partners and resulting hybrids	Example of junction alignment at homology boundary	Number of Events		
			DSBs	DSBs at repeats	
a NAHR			4	4	4
b NHEJ					
1					
2			49	98	42
3					
R1, R1' - similar repeats of the same class; R2 - a different repeat DR1, DR2 - direct repeats flanking Alu insertions			Total	53	102

Figure 4 | Most subtelomeric homology breakpoints are consistent with NHEJ. **a, b**, For each mechanistic scenario, we show both original and derived forms, assuming reciprocal exchange. One derived form would be lacking in non-reciprocal cases. The third column gives a schematic example of each scenario identified in pairwise alignments of subtelomeric homology blocks. Of the complete, non-redundant set of 56 homology breakpoints, 53 were assigned a mechanistic scenario (details in Supplementary Table S5 and Fig. S3). In some cases, two originals and one hybrid were available for comparison (for example, NHEJ group 1). Other predicted states were not among surviving, sequenced alleles.

using this approach were confounded by the use of probes encompassing several subtelomeric blocks, each with a different chromosomal distribution and evolutionary history^{5,26}. To refine the analysis of structural changes in subtelomeres, we used four small FISH probes, each of which encompasses a single homology block. This approach reveals an unanticipated degree of recent genomic rearrangement in subtelomeres. Each block varies in copy number and chromosomal location between human individuals (Fig. 6), and FISH detects more chromosomal sites than are evident in the genome assembly or hybrid panel (Supplementary Table S8). We detect content variation at 14 chromosomal ends using just four block probes on three individuals. Further analyses would undoubtedly uncover more variation.

Gross structural polymorphism of human subtelomeres is also evident in the finished sequence of allelic pairs (Fig. 1). The two sequenced alleles of 16p are 99.8% identical in chromosome-specific DNA sequence, transition to much lower identity ($\sim 93\%$) within the adjoining block 17, and have no detectable homology in distal sequence (Supplementary Fig. S6). The 19p alleles also differ grossly in subtelomeric content (Fig. 1). One of the structurally variant 4q alleles (4qA) is found in association with facioscapulohumeral dystrophy²⁷. Other cases of gross allelic variation are revealed by PCR analyses of the hybrid panel (Supplementary Fig. S1).

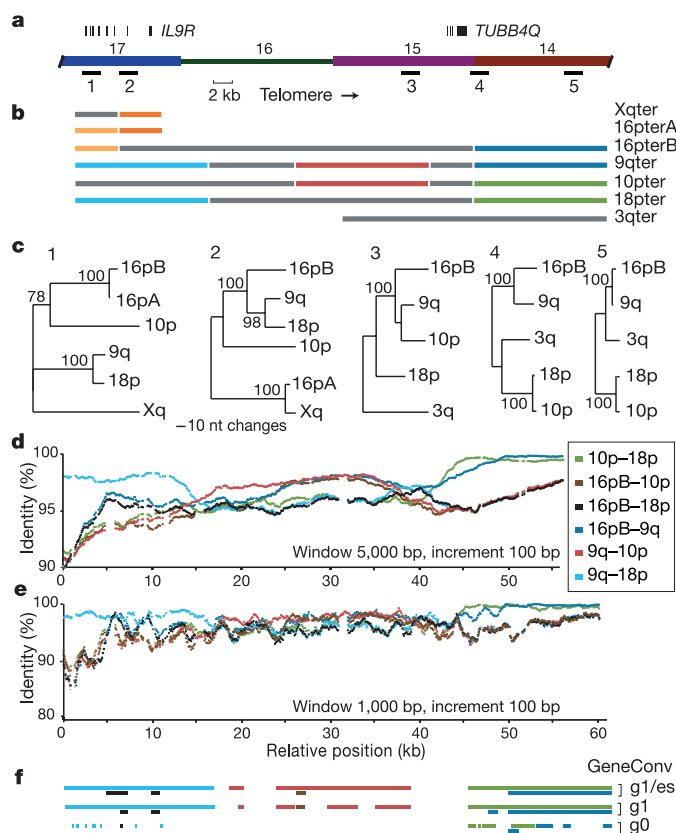


Figure 5 | Homology-based sequence transfers between subtelomeres. **a**, The region analysed encompasses four numbered blocks, two multi-exon genes, and five sequences sampled for phylogenetic analyses. **b**, Diagram of multiple sequence alignment with colours (excluding grey) indicating the best matching pairs with $\geq 98\%$ identity in non-overlapping 5-kb windows. **c**, Neighbour-joining trees with bootstrap values (over 1,000 replicates) constructed from 2-kb samples of the alignments. **d, e**, Plot of per cent identity between four subtelomeres in 5-kb (**d**) and 1-kb (**e**) windows. Colours indicate alignments of different pairs. **f**, The same colours indicate transferred segments found statistically significant by GeneConv using different stringency parameters.

Subtelomeric dynamics are not confined to the human lineage. Blocks moved, and copies were lost and gained during primate evolution (Fig. 6). For example, block 20 is present at ≥ 9 subtelomeric locations in chimpanzee and human, whereas it occurs at only a few internal sites in gorilla and orangutan. The odorant-receptor-gene-containing block 5 was completely lost from the orangutan and gorilla genomes, yet is duplicated in chimpanzee and human genomes in four or more sites. The high similarity between sequenced human copies of these blocks (Fig. 1 and Supplementary Table S2), together with the fact that humans have more copies of three of the four blocks than other primates, argues that the diversity of these particular block distributions arose primarily by recent duplications, rather than by loss of different subsets of ancestral copies. We estimate that an average of 25 independent events, involving relocation or copy-number change of a total of ~ 1.2 Mb, can account for the observed differences between chimpanzee and human in the subtelomeric distribution of these blocks. FISH analyses also suggest that chimpanzees, like humans, show gross variation in subtelomeric content. Future, less anthropocentric analyses will probably reveal subtelomeric blocks that have been lost in humans but retained and perhaps duplicated in other primates.

Timing and rates of subtelomeric transfers

The very recent nature of the interchromosomal events shaping subtelomeres is apparent from the high similarity of paralogous blocks on different chromosomes and from our cytogenetic analyses. For 28 of the 41 blocks, even the most dissimilar copies exceed 97% identity (Fig. 1 and Supplementary Table S2). Assuming a mutation rate of 10^{-3} substitutions per site per million years²⁸ (Myr), all the copies of these 28 blocks must have formed by duplication (except the original one) or participated in homology-based sequence transfer in the past 15 Myr during the divergence of humans and great apes.

Chromosome termini	Block 20				Block 5				Block 3				Block 2			
	HS1	HS2	HS3	PTR	GGO	PPY	HS1	HS2	HS3	PTR	GGO	PPY	HS1	HS2	HS3	PTR
1p																
2q																
3p																
3q																
4q																
5q																
6p																
6q																
7p																
8p																
8q																
9p																
9q																
11p																
12p																
15q																
16p																
16q																
17p																
19p																
19q																
20p																
20q																
Xq +/or 7q																
Internal sites																
1q41																
1q43																
2q14.1																
4q24-28																
4q32-33																
20q11.2																
Ycen																

Figure 6 | Chromosomal distribution of four subtelomeric blocks. FISH was conducted on three unrelated humans (HS1–3), chimpanzee (PTR), gorilla (GGO) and orangutan (PPY) (see Supplementary Methods). Coloured bars indicate sites at which FISH signals were consistently observed on both homologues (two bars) or only one homologue (one bar). Colours correspond to Fig. 1. Chromosome locations are given according to the human karyotype. No signal was observed for block 5 in gorilla and orangutan; its presence was also not detected by PCR (Supplementary Table S3). NA, not applicable.

When all pairs of human subtelomeric blocks are compared, the vast majority have 99.0–99.9% identity (Supplementary Fig. S7). Pairwise comparisons of all interchromosomal segmental duplications in the genome peak at $\sim 98\%$ identity (ref. 1), indicating that most subtelomeric segmental duplications result from more recent events than other segmental duplications. Indeed, after correcting for redundancy, we find that subtelomeres account for 40% of all duplications in the latest genome assembly³ that have a match of $\geq 98.7\%$ identity on another chromosome. Remarkably, ~ 1 Mb (40%) of known subtelomeric terrain has a paralogous match of $\geq 99.5\%$ identity, often rivaling the similarity of allelic copies.

We estimate conservatively that 49% (1.13 Mb) of known subtelomeric sequence was generated after humans and chimpanzees diverged (Supplementary Fig. S8). This amount equates to an observed rate of subtelomeric interchromosomal sequence duplication and/or transfer during the last 6.5 Myr of ~ 0.075 bases per site per Myr (Supplementary Table S9). We estimate from our cytogenetic analyses of the four blocks in Fig. 6 that these subtelomeric sequences relocated or changed in copy number at a rate of ~ 0.09 bases per site per Myr during the same time interval. The sequence- and cytogenetic-based estimation methods capture slightly different aspects of subtelomeric dynamics and underestimate the true rates of interchromosomal sequence transfer. Nevertheless, both estimates yield rates >60 -fold higher than those of point mutations²⁸ or bases added by retrotransposon insertion²⁹ over the same evolutionary period.

Given the amount of new subtelomeric sequence apparently created during the past 5 Myr (1.0 Mb), we estimate that ~ 7 gene duplicates arose in human subtelomeres per Myr in recent times (Supplementary Table S9). Even if half of these genes are deceptively young owing to sequence transfers between pre-existing copies, the rate of gene duplication in subtelomeres (0.04 duplicates per gene per Myr) is fourfold higher than the genome-wide average³⁰. The rate of gene creation in subtelomeres is matched only by that in pericentromeric regions, which, like subtelomeres, are hotbeds of segmental duplications³¹.

Discussion

Here we demonstrate that a multitude of predominantly NHEJ-mediated translocations led to a complex patchwork of segmental duplications in human subtelomeres that exchange sequence at a remarkably high rate. The extraordinary recent dynamics of subtelomeres complicate the description of the human genomic landscape and its variation. Perhaps no chromosome-specific marker or block organization exists within subtelomeres, as they appear to evolve as a pool of variant allelic and paralogous structures. More over, inter-allelic subtelomeric recombination rates may be impossible to quantify owing to the high frequency of interchromosomal transfers.

Why are subtelomeres so plastic? Deviations in copy number of subtelomeric DNA might be better tolerated than segmental aneuploidy of other genomic regions. Furthermore, subtelomeres might be more susceptible to DSBs and/or more readily repaired through interchromosomal interactions than other regions³². Telomere clustering in meiotic cells³³ might favour exchange of chromosome ends during DSB healing. Gross allelic differences probably make some subtelomeres prone to mispairing at meiosis, catalysing further change.

Subtelomeric rearrangements might not be restricted to the germline, but could also arise in somatic cells during repair of DSBs or eroded telomeres. The resulting genotypic heterogeneity might affect fitness at the cellular and/or individual levels. Indeed, subtelomeres coalesce with telomeres in DNA-repair foci in naturally senescent cells^{34,35} and in cells with artificially induced telomere dysfunction³⁶. Furthermore, the high level of apparent sister chromatid exchanges observed at chromosome ends (10^{-2} per Mb per generation)³⁷ signals a high DSB rate and could subsume interchromosomal subtelomeric exchanges.

The results of ectopic repair of subtelomeres could have advantages beyond the healing of damaged ends. With their propensity to duplicate and exchange, subtelomeres could serve as a nursery for new genes and a place where haplotypes can diversify faster than in single-copy genomic regions. Sequence transfer between paralogous genes (as in Fig. 5) has the potential to create advantageous new combinations of sequence variants, aiding adaptive peak shifts³⁸. Indeed, subtelomeric genes are associated with adaptive processes in other organisms (refs 39–41 and citations in ref. 5). However, subtelomeric dynamics are a double-edged sword. Some DSB repair events could result in loss or gain of dosage-sensitive genes in the most distal single-copy DNA, or in contextual changes with adverse effects on gene regulation. The sequence analyses presented here contribute to a developing framework that will enable exploration of the roles of subtelomeric dynamics in normal variation, adaptive change and clinically manifested disorders.

The translocation-based model developed here to explain subtelomeric segmental duplications could be broadly applicable to other interchromosomal segmental duplications (Fig. 2). The first step in this model is a reciprocal translocation (Fig. 2a), which arises *de novo* in $\sim 1/2,000$ concepti⁴². One in 500 healthy individuals carries a cytogenetically visible, balanced translocation⁴³. A second interchromosomal exchange between the translocation derivatives (Fig. 2b) is likely to be selectively favoured if it reduces the risk of passing a grossly imbalanced chromosomal complement to gametes. Duplicated segments, particularly when present on just one allele, can in turn promote translocations through NAHR (Fig. 2d). Furthermore, a DSB occurring in a hemizygous region stemming from an unbalanced translocation has increased probability of causing another translocation, inversion or intrachromosomal deletion, owing to the absence of a homologous template for its repair (Fig. 2e). Thus, segmental polymorphisms predispose to further rearrangements, which in turn lead to new segmentally polymorphic structures. This cycle of segmental polymorphism and gross genomic rearrangement is particularly obvious in subtelomeres and could underlie structural variation^{44–46} and genomic disorders⁴ arising at many other locations in the human genome.

METHODS

Additional results and methodological details, including the basis for all rate calculations, are provided as Supplementary Methods.

Sequence collation and analysis. Details of the iterative search for finished subtelomeric sequences are provided in the Supplementary Methods. Sequences with continuous overlap of >99.8% nucleotide identity were merged into contigs (Supplementary Table S1) and assumed to represent the same genomic region or an allelic variant. We used a combination of approaches to establish or verify the chromosome location of contigs (Supplementary Table S1), including PCR of a monochromosomal hybrid panel (Supplementary Table S3), FISH (Supplementary Table S10), and matches to half-YAC (yeast artificial chromosome) vector-insert junction sequences¹⁰ (Supplementary Table S11). Regions of similarity were identified from pairwise sequence alignments made by BLAST2⁴⁷, without masking repeats. Blocks of paralogy were delineated when one or more contigs showed a break in homology, except where paralogy adjoined a gap in available sequence. Block colour/number are changed in Fig. 1 if similarity is lost on one or more subtelomeres, except when loss of homology occurs within 3 kb of another breakpoint. However, all breakpoints were evaluated for mechanistic signatures (see below). Blocks from different chromosomal contigs were aligned using cross_match (<http://www.phrap.org/>) and MAVID⁴⁸. Per cent identities of block copies were calculated without insertions or deletions and with Jukes-Cantor correction for multiple substitutions. From 1,438 alignments (26.8 Mb total aligned sequence), a best-matching partner was identified for each block in each chromosomal contig (Supplementary Fig. S7). To remove redundancy in cases of reciprocal best matches, only one of the two alignments was included in the estimation of the amount of recently generated sequence (see Supplementary Methods and Supplementary Table S9). We also calculated the sum of non-overlapping interchromosomally duplicated bases with paralogous match of $\geq 98.7\%$ in subtelomeres or elsewhere in the latest genome assembly (Build 35), as outlined in Supplementary Methods.

Subtelomeric block analysis by PCR and FISH. The subtelomeric content of 24 individual human chromosomes isolated in a hybrid panel was analysed by PCR

using 160 primer pairs (Supplementary Fig. S1 and Table S3). FISH was performed as detailed in the Supplementary Methods, using block-specific probes generated by long-range PCR (blocks 20, 5 and 2) or cosmid f7501 (block 3; ref. 6) on primary cultures of three unrelated Caucasians (2 males and one female) and cell lines of male chimpanzee, orangutan and gorilla. The assumptions used to conservatively estimate the rate with which these blocks changed copy number or location since the divergence of humans and chimpanzees are given in Supplementary Methods. Note that this rate excludes homology-based sequence transfers among pre-existing copies, whereas the sequence-based estimate includes duplications and homology-based sequence transfers, but not changes in segment location.

Breakpoint analyses. We identified homology breakpoints from all pairwise subtelomeric sequence alignments and evaluated a nonredundant set for mechanistic signatures as described in the Supplementary Methods (Supplementary Table S5). All remaining block junctions in Fig. 1 are nearly identical replicas of members of this junction set, owing to their duplication within larger segments. The number of independent DSBs was counted as two for each deduced NHEJ event and one for each NAHR event in the non-redundant set. We queried the human genome by BLAT with the 200 bp surrounding each NHEJ breakpoint lacking a gene or known repeat, and we found no novel repeats.

Detection of homology-based transfer. Changes in per cent identity along pairwise sequence alignments were determined using the percentIDplot program (E. M. W. and E. V. L., unpublished data). The best-matching pair in each 5-kb and 2-kb window in each sequence was identified from a multiple sequence alignment generated using MAVID⁴⁸ (Supplementary Fig. S5). Phylogenetic trees were constructed using PAUP⁴⁹.

Received 7 April; accepted 5 July 2005.

- Samonte, R. V. & Eichler, E. E. Segmental duplications and the evolution of the primate genome. *Nature Rev. Genet.* **3**, 65–72 (2002).
- Bailey, J. A. *et al.* Recent segmental duplications in the human genome. *Science* **297**, 1003–1007 (2002).
- Bailey, J. A., Yavor, A. M., Massa, H. F., Trask, B. J. & Eichler, E. E. Segmental duplications: organization and impact within the current human genome project assembly. *Genome Res.* **11**, 1005–1017 (2001).
- Shaw, C. J. & Lupski, J. R. Implications of human genome architecture for rearrangement-based disorders: the genomic basis of disease. *Hum. Mol. Genet.* **13** (review issue 1), R57–R64 (2004).
- Mefford, H. & Trask, B. J. The complex structure and dynamic evolution of human subtelomeres. *Nature Rev. Genet.* **3**, 91–102 (2002).
- Trask, B. J. *et al.* Members of the olfactory receptor gene family are contained in large blocks of DNA duplicated polymorphically near the ends of human chromosomes. *Hum. Mol. Genet.* **7**, 13–26 (1998).
- Monfouilloux, S. *et al.* Recent human-specific spreading of a subtelomeric domain. *Genomics* **51**, 165–176 (1998).
- Martin, C. L. *et al.* The evolutionary origin of human subtelomeric homologies—or where the ends begin. *Am. J. Hum. Genet.* **70**, 972–984 (2002).
- Fan, Y., Linardopoulou, E., Friedman, C., Williams, E. M. & Trask, B. J. Genomic structure and evolution of the ancestral chromosome fusion site in 2q13–2q14.1. *Genome Res.* **12**, 1651–1662 (2002).
- Riethman, H. C. *et al.* Integration of telomere sequences with the draft human genome sequence. *Nature* **409**, 948–951 (2001).
- Linardopoulou, E. *et al.* Transcriptional activity of multiple copies of a subtelomerically located olfactory receptor gene that is polymorphic in number and location. *Hum. Mol. Genet.* **10**, 2373–2383 (2001).
- Knight, S. J. & Flint, J. The use of subtelomeric probes to study mental retardation. *Methods Cell Biol.* **75**, 799–831 (2004).
- Brown, W. R. *et al.* Structure and polymorphism of human telomere-associated DNA. *Cell* **63**, 119–132 (1990).
- de Lange, T. *et al.* Structure and variability of human chromosome ends. *Mol. Cell Biol.* **10**, 518–527 (1990).
- Flint, J. *et al.* Sequence comparison of human and yeast telomeres identifies structurally distinct subtelomeric domains. *Hum. Mol. Genet.* **6**, 1305–1313 (1997).
- Riethman, H. *et al.* Mapping and initial analysis of human subtelomeric sequence assemblies. *Genome Res.* **14**, 18–28 (2004).
- Smit, A. F. & Riggs, A. D. Tiggers and DNA transposon fossils in the human genome. *Proc. Natl Acad. Sci. USA* **93**, 1443–1448 (1996).
- Eichler, E. E., Archidiacono, N. & Rocchi, M. CAGGG repeats and the pericentromeric duplication of the hominoid genome. *Genome Res.* **9**, 1048–1058 (1999).
- Pfeiffer, P., Goedecke, W. & Obe, G. Mechanisms of DNA double-strand break repair and their potential to induce chromosomal aberrations. *Mutagenesis* **15**, 289–302 (2000).
- Rothkamm, K., Kruger, I., Thompson, L. H. & Lobrich, M. Pathways of DNA double-strand break repair during the mammalian cell cycle. *Mol. Cell Biol.* **23**, 5706–5715 (2003).
- Bailey, J. A., Liu, G. & Eichler, E. E. An Alu transposition model for the origin

- and expansion of human segmental duplications. *Am. J. Hum. Genet.* **73**, 823–834 (2003).
22. Zhou, Y. & Mishra, B. Quantifying the mechanisms for segmental duplications in mammalian genomes by statistical analysis and modeling. *Proc. Natl Acad. Sci. USA* **102**, 4151–4156 (2005).
 23. Nergadze, S. G., Rocchi, M., Azzalin, C. M., Mondello, C. & Giulotto, E. Insertion of telomeric repeats at intrachromosomal break sites during primate evolution. *Genome Res.* **14**, 1704–1710 (2004).
 24. Sawyer, S. Statistical tests for detecting gene conversion. *Mol. Biol. Evol.* **6**, 526–538 (1989).
 25. Mefford, H. C., Linardopoulou, E., Coil, D., van den Engh, G. & Trask, B. J. Comparative sequencing of a multicopy subtelomeric region containing olfactory receptor genes reveals multiple interactions between non-homologous chromosomes. *Hum. Mol. Genet.* **10**, 2363–2372 (2001).
 26. Der-Sarkissian, H., Vergnaud, G., Borde, Y. M., Thomas, G. & Londono-Vallejo, J. A. Segmental polymorphisms in the proterminal regions of a subset of human chromosomes. *Genome Res.* **12**, 1673–1678 (2002).
 27. Lemmers, R. J. *et al.* Facioscapulohumeral muscular dystrophy is uniquely associated with one of the two variants of the 4q subtelomere. *Nature Genet.* **32**, 235–236 (2002).
 28. Chen, F. C. & Li, W. H. Genomic divergences between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees. *Am. J. Hum. Genet.* **68**, 444–456 (2001).
 29. Liu, G. *et al.* Analysis of primate genomic variation reveals a repeat-driven expansion of the human genome. *Genome Res.* **13**, 358–368 (2003).
 30. Lynch, M. & Conery, J. S. The evolutionary fate and consequences of duplicate genes. *Science* **290**, 1151–1155 (2000).
 31. She, X. *et al.* The structure and evolution of centromeric transition regions within the human genome. *Nature* **430**, 857–864 (2004).
 32. Ricchetti, M., Dujon, B. & Fairhead, C. Distance from the chromosome end determines the efficiency of double strand break repair in subtelomeres of haploid yeast. *J. Mol. Biol.* **328**, 847–862 (2003).
 33. Bass, H. W. Telomere dynamics unique to meiotic prophase: formation and significance of the bouquet. *Cell. Mol. Life Sci.* **60**, 2319–2324 (2003).
 34. d'Adda di Fagagna, F. *et al.* A DNA damage checkpoint response in telomere-initiated senescence. *Nature* **426**, 194–198 (2003).
 35. Zou, Y., Sfeir, A., Gryaznov, S. M., Shay, J. W. & Wright, W. E. Does a sentinel or a subset of short telomeres determine replicative senescence? *Mol. Biol. Cell* **15**, 3709–3718 (2004).
 36. Takai, H., Smogorzewska, A. & de Lange, T. DNA damage foci at dysfunctional telomeres. *Curr. Biol.* **13**, 1549–1556 (2003).
 37. Cornforth, M. N. & Eberle, R. L. Termini of human chromosomes display elevated rates of mitotic recombination. *Mutagenesis* **16**, 85–89 (2001).
 38. Hansen, T. F., Carter, A. J. & Chiu, C. H. Gene conversion may aid adaptive peak shifts. *J. Theor. Biol.* **207**, 495–511 (2000).
 39. Halme, A., Bumgarner, S., Styles, C. & Fink, G. R. Genetic and epigenetic regulation of the *FLO* gene family generates cell-surface variation in yeast. *Cell* **116**, 405–415 (2004).
 40. Fabre, E. *et al.* Comparative genomics in hemiascomycete yeasts: evolution of sex, silencing, and subtelomeres. *Mol. Biol. Evol.* **22**, 856–873 (2005).
 41. De Las Penas, A. *et al.* Virulence-related surface glycoproteins in the yeast pathogen *Candida glabrata* are encoded in subtelomeric clusters and subject to *RAP1*- and *SIR*-dependent transcriptional silencing. *Genes Dev.* **17**, 2245–2258 (2003).
 42. Warburton, D. *De novo* balanced chromosome rearrangements and extra marker chromosomes identified at prenatal diagnosis: clinical significance and distribution of breakpoints. *Am. J. Hum. Genet.* **49**, 995–1013 (1991).
 43. Genetics and Public Policy Center. Genetics Information: Translocations. (<http://www.dnapolicy.org/genetics/translocations.jhtml>) (2004).
 44. Wong, Z., Royle, N. J. & Jeffreys, A. J. A novel human DNA polymorphism resulting from transfer of DNA from chromosome 6 to chromosome 16. *Genomics* **7**, 222–234 (1990).
 45. Sebat, J. *et al.* Large-scale copy number polymorphism in the human genome. *Science* **305**, 525–528 (2004).
 46. lafrate, A. J. *et al.* Detection of large-scale variation in the human genome. *Nature Genet.* **36**, 949–951 (2004).
 47. Tatusova, T. A. & Madden, T. L. BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences. *FEMS Microbiol. Lett.* **174**, 247–250 (1999).
 48. Bray, N. & Pachter, L. MAVID multiple alignment server. *Nucleic Acids Res.* **31**, 3525–3526 (2003).
 49. Swofford, D. L. *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods)* (Sinauer Associates, Sunderland, 2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to the many contributors to the Human Genome Project who generated the sequences that made this study possible, and to the Eichler and Haussler laboratory groups for making data on segmental duplications readily accessible. Our work was supported by the NIH. We thank E. Eichler, H. Malik, D. Gottschling, J. Gogarten, K. Rudd and M. Schlador for comments on the manuscript, and J. Felsenstein for advice on phylogenetic analyses.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to B.J.T. (btrask@fhcrc.org).

Conservation of Y-linked genes during human evolution revealed by comparative sequencing in chimpanzee

Jennifer F. Hughes¹, Helen Skaletsky¹, Tatyana Pyntikova¹, Patrick J. Minx², Tina Graves², Steve Rozen¹, Richard K. Wilson² & David C. Page¹

The human Y chromosome, transmitted clonally through males, contains far fewer genes than the sexually recombining autosome from which it evolved. The enormity of this evolutionary decline has led to predictions that the Y chromosome will be completely bereft of functional genes within ten million years^{1,2}. Although recent evidence of gene conversion within massive Y-linked palindromes runs counter to this hypothesis, most unique Y-linked genes are not situated in palindromes and have no gene conversion partners^{3,4}. The 'impending demise' hypothesis thus rests on understanding the degree of conservation of these genes. Here we find, by systematically comparing the DNA sequences of unique, Y-linked genes in chimpanzee and human, which diverged about six million years ago, evidence that in the human lineage, all such genes were conserved through purifying selection. In the chimpanzee lineage, by contrast, several genes have sustained inactivating mutations. Gene decay in the chimpanzee lineage might be a consequence of positive selection focused elsewhere on the Y chromosome and driven by sperm competition.

The human X and Y chromosomes co-evolved from an ordinary pair of autosomes that existed in the mammalian ancestor roughly 300 million years ago⁵. Because most of the Y chromosome does not participate in sexual recombination, it has degenerated substantially, both in size and gene content, in comparison with the X chromosome⁶. Recent studies of the ampliconic region of the Y chromosome, which comprises almost half of the chromosome's euchromatin, revealed large palindromes where abundant gene conversion may forestall gene decay^{3,4}. However, nearly all of the remainder of the Y chromosome's genes are found in the X-degenerate regions, which were once identical in sequence to the X chromosome but have since diverged substantially³. Unlike the ampliconic sequence, the X-degenerate sequence does not routinely undergo recombination of any sort, so rapid, ongoing gene loss might be expected there.

To understand better the recent evolution of the human X-degenerate sequence and the fate of its remaining genes, we determined the nucleotide sequence of the X-degenerate portion of the chimpanzee Y chromosome. The resulting sequence spans 9.5 megabases (Mb), is complete apart from two small gaps, and is accurate to about one nucleotide per 200,000.

Before using these sequence data to test the impending demise hypothesis, we compared basic characteristics of the X-degenerate sequences in chimpanzee and human, discovering that both the gross structures and nucleotide sequences of hominoid Y chromosomes have evolved rapidly. Counterpoint to the Y chromosome's rapid evolution is provided by human chromosome 21 and its orthologue, chimpanzee chromosome 22, the only autosomes fully sequenced in

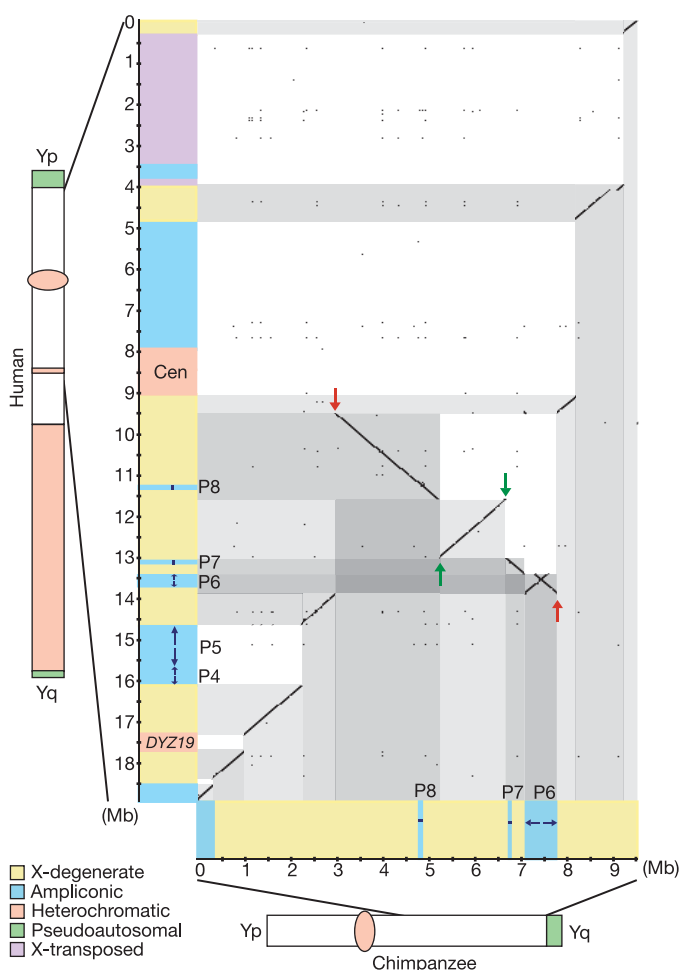


Figure 1 | Dot-plot comparison of the X-degenerate region of the chimpanzee Y chromosome (below) with the euchromatic region of the human Y chromosome (left). These chromosomal regions are shown schematically on the axes, where major features, including palindromes P4–P8, are indicated. Each dot within the plot represents 100% identity within a 200-bp window. Within the plot, grey shading indicates blocks of uninterrupted sequence alignment. Break points of two inversions (1.5 and 5.0 Mb) are indicated by arrows (green and red, respectively). Supplementary Fig. 1 provides a more highly annotated version of this plot, including gene and pseudogene positions. Cen, centromere.

¹Howard Hughes Medical Institute, Whitehead Institute, and Department of Biology, Massachusetts Institute of Technology, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA. ²Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Boulevard, St Louis, Missouri 63108, USA.

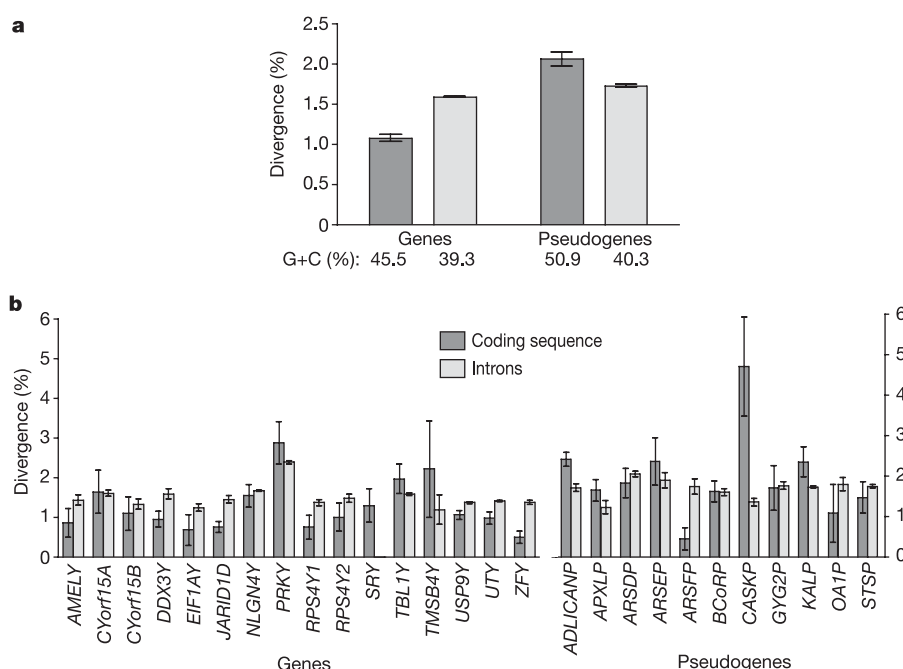


Figure 2 | Human-chimpanzee divergence in coding sequence and introns of X-degenerate genes and pseudogenes. **a**, Aggregate per cent divergences for coding and intron sequences of all genes and all pseudogenes. Listed below are G + C contents for each sequence class. See Supplementary Table 3 for numeric data and measures of statistical significance. **b**, Left: graph of coding and intron divergences for each gene, analysed separately and listed alphabetically. Right: an analogous graph of divergences for each pseudogene. Error bars depict standard errors for uncorrected per cent divergence calculated from 500 bootstrap replicates using MEGA2 software²⁴.

both species. Whereas the nucleotide sequences of human chromosome 21 and chimpanzee chromosome 22 are grossly co-linear⁷, implying that little structural change has occurred, the Y chromosome has undergone significant restructuring since the chimpanzee and human lineages diverged (Fig. 1; see also Supplementary Fig. 1). In humans, the X-degenerate sequences are distributed along both arms of the Y chromosome, and they are interrupted at several points by large blocks of ampliconic, heterochromatic, or other sequences. In the chimpanzee, by contrast, the X-degenerate sequences are found in a single, nearly contiguous block on the long arm of the Y chromosome. In addition, the chimpanzee and human X-degenerate sequences differ by two large inversions (Supplementary Figs 2–4). The larger inversion, spanning nearly 5 Mb, occurred in the chimpanzee lineage. The smaller, 1.5-Mb inversion occurred in the human lineage. Such inversions may be of relatively little consequence in the male-specific region of the Y chromosome, which does not engage in crossing-over with a chromosomal homologue. (In heterozygotes for autosomal or X-linked inversions, crossing-over within the inverted segment results in genetically unbalanced offspring.) The nucleotide sequences of human chromosome 21 and chimpanzee chromosome 22 have been reported to display 1.44% divergence⁷. We observed significantly greater nucleotide divergence, 1.72%, between the X-degenerate sequences of the chimpanzee and human Y chromosomes. This difference is unsurprising given previous evidence of accelerated DNA sequence evolution on the Y chromosome, the result of its being transmitted exclusively through the substitution-prone male germ line⁸.

The density of interspersed repetitive elements is nearly identical in the X-degenerate regions of the chimpanzee and human Y chromosomes—at least when one combines all such repeat classes (Supplementary Table 1). However, we found evidence of marked differences between the chimpanzee and human lineages in the levels of transposition activity of the three main classes of retroelements (Supplementary Table 2). We found that Alu elements were more active in the human lineage, whereas long interspersed nucleotide (L1) elements and especially endogenous retroviruses were more active in the chimpanzee lineage. Most notably, the chimpanzee sequence contains 21 copies of two novel endogenous retroviruses, CERV1 and CERV2, which are completely absent from the human genome.

Having completed these basic comparisons of the chimpanzee and human X-degenerate sequences, we then addressed a prediction of the impending demise model. The model's central premise is that, in

recent times, the human Y chromosome has been losing genes once shared with the X chromosome at a pace approximating 5 genes per million years^{1,2}. Assuming that Y-linked gene decay and loss occurred randomly, and that the chimpanzee and human lineages have been separate for about six million years, the chimpanzee Y chromosome should carry many genes that have no functional orthologue on the human Y chromosome.

To test this prediction, we characterized the gene content of the chimpanzee X-degenerate sequence by several means. First, we electronically searched the sequence for orthologues of all known human X-degenerate genes and pseudogenes. (The human pseudogenes, which do not seem to be transcribed, bear inactivating mutations that disrupt or delete splice sites and exons, or that interrupt or shift open reading frames³.) We identified chimpanzee orthologues of all 16 such genes and all 11 such pseudogenes. Notably, the chimpanzee counterparts of the 11 human pseudogenes are also pseudogenes, with the great majority of inactivating mutations being shared between the two species, indicating that all 11 pseudogenes were inactivated before divergence of the chimpanzee and human lineages. This suggests that none of the 11 human X-degenerate pseudogenes has lost its functionality during the last six million years of human evolution. In addition, we conducted GenScan⁹ and BLAST^{10,11} searches of the chimpanzee X-degenerate sequence for transcription units that have no human Y counterpart. We found no such chimpanzee-specific transcription units. Thus, comparative cataloguing of X-degenerate genes and pseudogenes in the chimpanzee and human suggests that little or no X-degenerate gene loss or decay has occurred during the last six million years of human evolution. These findings contradict the model of the human Y chromosome's impending demise, and instead provide empirical support for mathematical models of sex chromosome evolution that predict a slowing of the rate of gene decay as Y chromosomes evolve⁶.

These findings also suggest that purifying selection on the Y chromosome has been more effective during recent human evolution than previously supposed. To examine this hypothesis, we compared the degree of human–chimpanzee divergence in the X-degenerate genes' coding regions with the degree of divergence in their introns, which served as controls. As additional controls, we examined interspecies divergence in the X-degenerate pseudogenes. Genes for which the protein products are subject to purifying selection should exhibit less interspecies divergence in coding sequences than in introns. We found this to be the case for the X-degenerate genes as

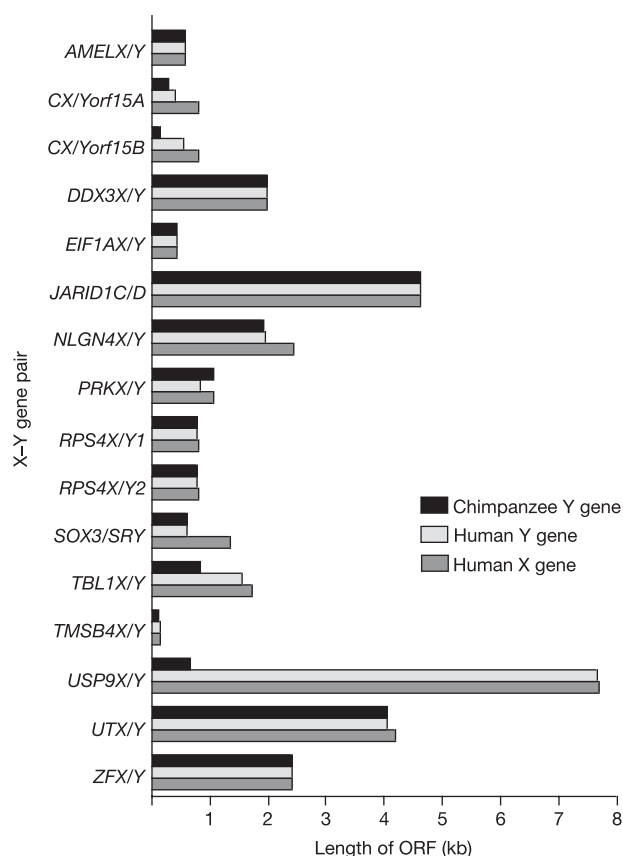


Figure 3 | Lengths of coding sequences of X-degenerate genes on chimpanzee and human Y chromosomes, and their human X-linked homologues. For each X–Y gene pair, three horizontal bars are shown, representing the predicted lengths (in kb) of the coding regions for the chimpanzee Y gene (black), its human Y-linked orthologue (light grey) and its human X-linked homologue (dark grey). Genes are listed in alphabetical order.

a group ($P < 0.00001$; Fig. 2a; see also Supplementary Table 3). As expected, this was not true for the X-degenerate pseudogenes, where human–chimpanzee divergence proved to be similar or even greater in (former) coding sequences than in (former) introns. (The elevated G + C content of these former coding sequences probably accounts for the elevated rate of sequence evolution¹².) When every gene was analysed independently, most exhibited the same trend towards greater

conservation of coding sequence relative to introns (Fig. 2b). We conclude that purifying selection has been a potent force in maintaining X-degenerate gene function during recent human evolution.

Whereas the repertoire of X-degenerate genes present in the human/chimpanzee ancestor evidently remained functionally intact during subsequent human evolution, we discovered evidence of significant gene decay during chimpanzee evolution. We used three types of analysis: coding sequence divergence (Fig. 2), open reading frame (ORF) integrity (Fig. 3), and transcriptional activity (as assayed by polymerase chain reaction with reverse transcription (RT–PCR); Supplementary Fig. 6). Interspecies divergence of coding sequence differed substantially among the X-degenerate genes (Fig. 2), prompting us to rank all of the genes by this measure (Table 1). We reasoned that genes displaying relatively high interspecies divergence might have been subject to relaxed selective constraints, and possibly functional decay, in the chimpanzee or human lineage, or both. Of the eight X-degenerate genes displaying >1.0% divergence, five had ORFs that were substantially truncated in the chimpanzee Y chromosome as compared with the human Y chromosome (Fig. 3). These ORF truncations were due to point mutations that disrupted splice sites or introduced stop codons in the chimpanzee genes, but did not grossly alter their genomic size or structure (Table 1).

Intriguingly, the truncated ORFs in chimpanzee include orthologues of both the largest and smallest proteins predicted to be encoded by the human Y chromosome³. In human, *USP9Y* has a critical role in spermatogenesis¹³ and is predicted to encode a 2,555-amino-acid ubiquitin protease that is 91% identical across its length to a 2,563-amino-acid protein encoded by the X chromosome. In chimpanzee, by contrast, the longest ORF in *USP9Y* would encode a protein of only 675 amino acids, with no intact catalytic domain (Supplementary Fig. 7). In human, *TMSB4Y* is expressed throughout the body and is predicted to encode a 44-amino-acid peptide that differs by three amino substitutions from an otherwise identical peptide encoded by its mammalian X-linked homologue. In chimpanzee, we detected no evidence of *TMSB4Y* transcription in any tested tissue or cell line (Supplementary Fig. 6), and the single splice donor site within the coding region has been lost through mutation (Supplementary Fig. 8). These examples illustrate the diversity of Y-linked, X-degenerate gene functions that have decayed in chimpanzee but not human.

Why have X-degenerate genes decayed in the chimpanzee lineage but not in the human lineage? We speculate that X-degenerate gene decay in the chimpanzee lineage may be a by-product of strong positive selection focused elsewhere on the Y chromosome, through a process known as genetic hitchhiking. Because the Y chromosome does not participate in sexual recombination with a chromosome

Table 1 | Chimpanzee gene characterization

Gene	Nucleotide divergence from human (%)	ORF length (% of human)	ORF-disrupting mutations in chimpanzee*
<i>PRKY</i>	2.89	128†	–
<i>TMSB4Y</i>	2.22	42	Splice donor (GT → GC)
<i>TBL1Y</i>	1.98	52	Splice acceptor (AG → GG); splice donor (GT → AT)
<i>CYorf15A</i>	1.65	72	–1 frame shift
<i>NLGN4Y</i>	1.55	99	–
<i>SRY</i>	1.31‡	100	–
<i>CYorf15B</i>	1.10	26	Nonsense (AAA → TAA)
<i>USP9Y</i>	1.07	9	Three splice donors (4-bp deletion; GT → AT; GT → AT)
<i>RSP4Y2</i>	1.01	100	–
<i>UTY</i>	0.99	100	–
<i>DDX3Y</i>	0.96	100	–
<i>AMELY</i>	0.86	100	–
<i>JARID1D</i>	0.76	100	–
<i>RPS4Y1</i>	0.76	100	–
<i>EIF1AY</i>	0.69	100	–
<i>ZFY</i>	0.50	100	–

Genes are ordered according to per cent coding sequence divergence from their human orthologues, from highest to lowest.

*We inferred the lineage in which each mutation occurred by comparison to the human X homologue.

†The ORF of *PRKY* is longer in the chimpanzee because of a 55-bp genomic deletion, near the 3' end of the gene, that occurred in the human lineage.

‡The sex-determining gene *SRY* is subject to positive selection²², which probably accounts for its relatively high interspecies divergence.

homologue, natural selection acts on the chromosome as a unit. Deleterious mutations in some Y-linked genes can be carried along, even to the point of fixation in a population, by physical linkage to strongly beneficial mutations in other Y-linked genes^{6,14}. In addition to their X-degenerate genes, primate Y chromosomes contain many families of ampliconic genes, which have testes-restricted expression patterns and critical functions in sperm production^{3,15}. Because of this central role in spermatogenesis, the Y chromosome's ampliconic genes may be subject to powerful selective pressures^{16–18}, especially in species such as chimpanzees where females usually mate with multiple males, the sperm of which then compete for a limited number of oocytes¹⁹. During chimpanzee evolution some X-degenerate genes may have been casualties of selective forces directed at the Y chromosome's ampliconic genes—forces that were not as intense during the evolution of our less promiscuous species. In the future, comparisons of sex chromosome variability²⁰ in chimpanzees and humans may provide a test of this speculative hypothesis.

METHODS

Mapping and sequencing. We mapped and sequenced a tiling path of 73 bacterial artificial chromosome (BAC) and 7 fosmid clones. Clones for sequencing were selected from two BAC libraries (CHORI-251, RPCI-43) and one fosmid library (CHORI-1251) (<http://bacpac.chori.org>). The CHORI-251 BAC and CHORI-1251 fosmid libraries originate from the same male chimpanzee. Only eight of the BAC clones in the tiling path are from the RPCI-43 library, which originates from a second male chimpanzee. We screened the BAC libraries with 11 pools of hybridization probes derived from 209 STS markers located within the X-degenerate region of the human Y chromosome^{3,21}. Additional probes and markers were obtained from chimpanzee BAC end sequences. Fosmid end sequences were used to identify appropriate clones for filling gaps in the BAC contigs. See Supplementary Fig. 5 for the complete clone contig map. The accuracy of the sequence was estimated using all available CHORI-251 BAC overlaps in the assembled tiling path. There were a total of 26 errors in over 5.31 Mb of aligned sequence, which correlates to 1 error per 204 kilobases (kb). Two gaps remain in the sequence and their sizes were estimated based on human sequence to be roughly 14 kb and 69 kb. These sizes were confirmed by fibre fluorescence *in situ* hybridization (FISH) analysis using a cell line derived from the same chimpanzee used in constructing the CHORI-251 and CHORI-1251 libraries (Supplementary Fig. 9).

Assessing the completeness of chimpanzee X-degenerate sequence coverage. We used the data set based on the 13 November 2003 assembly from the Chimpanzee Sequencing Consortium (http://www.ensembl.org/Pan_troglydotes) to search for the existence of chimpanzee X-degenerate sequence that is absent from the human sequence using two strategies. First, we used all chimpanzee unassigned contigs that were not identified as interspersed repeats (10,750 contigs) in a BLAST search of the human genome, presuming that those contigs that were a closest match to human X chromosome sequence but displayed less than 97% identity were candidate X-degenerate contigs. However, none was found, and instead all chimpanzee contigs that matched the human X sequence were 99% identical and are presumably from the X chromosome. Second, we used all known and predicted human X chromosome genes in a BLAST search of the chimpanzee unassigned contigs in an attempt to identify chimpanzee X-degenerate genes or pseudogenes that are not on the human Y chromosome. No significant matches were found.

Sequence alignment and dot-plot analysis. Chimpanzee and human sequences were aligned using Clustal W with default parameters²². Dot-plot analysis was performed using custom Perl code, which is available upon request.

Assigning insertions and deletions to the chimpanzee or human lineage. Insertions and deletions (indels) were identified by aligning the chimpanzee and human sequences. For indels that were at least 100 base pairs (bp) in length, we determined the nature of the mutation (insertion or deletion) and the lineage in which it occurred as follows. If the indel sequence consisted entirely of a known, full-length interspersed repeat, such as an Alu or L1, it was inferred to be the result of an insertion event. Because integrations of partial L1 elements are known to occur, if these were identified the corresponding sequence in the other species was examined to look for the presence of the same L1 element. This was done to avoid misclassifying partial L1 deletions as integrations of non-full-length elements. Indel sequences not classified as repetitive element insertions or tandem duplications were presumed to be the result of deletion events. Putative deleted sequences that were not composed entirely of interspersed repeats were used in a BLAST search of the non-redundant GenBank database to ensure that they were not transposed sequences, which would be evidenced by close matches

to an autosome. For the majority of these sequences (47 of 61), the second best match (after Y chromosome sequence) was the human X chromosome. This is expected because the X chromosome represents the ancestral state of the X-degenerate sequences. The remaining indels matched only chimpanzee or human Y chromosome sequence. These results were interpreted as corroborating a deletion event.

RT-PCR analysis. The RNeasy kit (Qiagen) was used to isolate total RNAs from chimpanzee male tissues (testis, liver, lung and spleen) and a chimpanzee male lymphoblastoid cell line; all tissues were obtained from Yerkes National Primate Research Center. RT-PCR primer sequences and product sizes are listed in Supplementary Table 4.

Received 15 April; accepted 3 August 2005.

1. Aitken, R. J. & Marshall Graves, J. A. The future of sex. *Nature* **415**, 963 (2002).
2. Graves, J. A. The degenerate Y chromosome—can conversion save it? *Reprod. Fertil. Dev.* **16**, 527–534 (2004).
3. Skaletsky, H. *et al.* The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* **423**, 825–837 (2003).
4. Rozen, S. *et al.* Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. *Nature* **423**, 873–876 (2003).
5. Lahn, B. T. & Page, D. C. Four evolutionary strata on the human X chromosome. *Science* **286**, 964–967 (1999).
6. Charlesworth, B. & Charlesworth, D. The degeneration of Y chromosomes. *Phil. Trans. R. Soc. Lond. B* **355**, 1563–1572 (2000).
7. Watanabe, H. *et al.* DNA sequence and comparative analysis of chimpanzee chromosome 22. *Nature* **429**, 382–388 (2004).
8. Miyata, T., Hayashida, H., Kuma, K., Mitsuyasu, K. & Yasunaga, T. Male-driven molecular evolution: a model and nucleotide sequence analysis. *Cold Spring Harb. Symp. Quant. Biol.* **52**, 863–867 (1987).
9. Yeh, R. F., Lim, L. P. & Burge, C. B. Computational inference of homologous gene structures in the human genome. *Genome Res.* **11**, 803–816 (2001).
10. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
11. Hellmann, I. *et al.* Selection on human genes as revealed by comparisons to chimpanzee cDNA. *Genome Res.* **13**, 831–837 (2003).
12. Casane, D., Boissinot, S., Chang, B. H., Shimmin, L. C. & Li, W. Mutation pattern variation among regions of the primate genome. *J. Mol. Evol.* **45**, 216–226 (1997).
13. Sun, C. *et al.* An azoospermic man with a *de novo* point mutation in the Y-chromosomal gene *USP9Y*. *Nature Genet.* **23**, 429–432 (1999).
14. Rice, W. R. Genetic hitchhiking and the evolution of reduced genetic activity of the Y sex chromosome. *Genetics* **116**, 161–167 (1987).
15. Lahn, B. T. & Page, D. C. Functional coherence of the human Y chromosome. *Science* **278**, 675–680 (1997).
16. Parker, G. A. Sperm competition and its evolutionary consequences in the insects. *Biol. Rev.* **45**, 525–567 (1970).
17. Dorus, S., Evans, P. D., Wyckoff, G. J., Choi, S. S. & Lahn, B. T. Rate of molecular evolution of the seminal protein gene *SEMG2* correlates with levels of female promiscuity. *Nature Genet.* **36**, 1326–1329 (2004).
18. Wyckoff, G. J., Wang, W. & Wu, C. I. Rapid evolution of male reproductive genes in the descent of man. *Nature* **403**, 304–309 (2000).
19. Dixon, A. F. *Primate Sexuality: Comparative Studies of the Prosimians, Monkeys, Apes and Human Beings* (Univ. Chicago Press, Chicago, 1998).
20. Filatov, D. A., Moneger, F., Negruțiu, I. & Charlesworth, D. Low variability in a Y-linked plant gene and its implications for Y-chromosome evolution. *Nature* **404**, 388–390 (2000).
21. Tilford, C. A. *et al.* A physical map of the human Y chromosome. *Nature* **409**, 943–945 (2001).
22. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).
23. Whitfield, L. S., Lovell-Badge, R. & Goodfellow, P. N. Rapid sequence evolution of the mammalian sex-determining gene *SRY*. *Nature* **364**, 713–715 (1993).
24. Kumar, S., Tamura, K., Jakobsen, I. B. & Nei, M. MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* **17**, 1244–1245 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by the National Institutes of Health and the Howard Hughes Medical Institute.

Author Information GenBank accession numbers for CERV1 and CERV2 are AY692036 and AY692037, respectively. GenBank accession numbers for all complementary DNA sequences are listed in Supplementary Table 5; accession numbers for all BAC and fosmid clones are listed in Supplementary Table 6. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.C.P. (page_admin@wi.mit.edu).

First fossil chimpanzee

Sally McBrearty¹ & Nina G. Jablonski²

There are thousands of fossils of hominins, but no fossil chimpanzee has yet been reported. The chimpanzee (*Pan*) is the closest living relative to humans¹. Chimpanzee populations today are confined to wooded West and central Africa, whereas most hominin fossil sites occur in the semi-arid East African Rift Valley. This situation has fuelled speculation regarding causes for the divergence of the human and chimpanzee lineages five to eight million years ago. Some investigators have invoked a shift from wooded to savannah vegetation in East Africa, driven by climate change, to explain the apparent separation between chimpanzee and human ancestral populations and the origin of the unique hominin locomotor adaptation, bipedalism^{2–5}. The Rift Valley itself functions as an obstacle to chimpanzee occupation in some scenarios⁶. Here we report the first fossil chimpanzee. These fossils, from the Kapthurin Formation, Kenya, show that representatives of *Pan* were present in the East African Rift Valley during the Middle Pleistocene, where they were contemporary with an extinct species of *Homo*. Habitats suitable for both hominins and chimpanzees were clearly present there during this period, and the Rift Valley did not present an impenetrable barrier to chimpanzee occupation.

The Kapthurin Formation forms the Middle Pleistocene portion of the Tugen Hills sequence west of Lake Baringo (Figs 1 and 2). It consists of a package of fluvial, lacustrine and volcanic sediments ~125 m thick, exposed over ~150 km² (refs 7–9) that contains numerous palaeontological and archaeological sites^{9–11}. It is divided into five members informally designated K1–K5 (ref. 7), and the sequence is well calibrated by ⁴⁰Ar/³⁹Ar dating¹².

Hominin fossils attributed to *Homo erectus* or *Homo rhodesiensis* have been found in the fluvial sediments of K3 (refs 11, 13, 14). The new chimpanzee fossils were discovered at Locality (Loc.) 99 in K3', the lacustrine facies of the same geological member. Loc. 99 consists of ~80 m² of exposures at an outcrop ~1 km northeast of site GnJh-19 where hominin mandible KNM-BK (Kenya National Museum-Baringo Kapthurin) 8518 was found¹⁴. Two chimpanzee fossils, KNM-TH (Kenya National Museum-Tugen Hills) 45519 and KNM-TH 45520, were found in surface context within an area of ~12 m² within Loc. 99; additional specimens (KNM-TH 45521 and KNM-TH 45522) were recovered from sieved superficial sediments within the same restricted area. The age of the chimpanzee fossils is constrained by ⁴⁰Ar/³⁹Ar dates of 545 ± 3 kyr (thousand years) on underlying K2 and 284 ± 12 kyr on overlying K4 (ref. 12). Because they are derived from a position low in this stratigraphic interval, they are probably closer to the maximum age of 545 kyr. *Homo* fossils KNM-BK 63-67 and KNM-BK 8518 from K3 are bracketed by ⁴⁰Ar/³⁹Ar dates of 543 ± 4 kyr and 509 ± 9 kyr¹² (Fig. 2).

K3' sediments are exposed in an outcrop of ~1 km² in the eastern portion of the Kapthurin Formation. They consist of black and red zeolitized clays interbedded with sands and heavily altered volcanics. Sedimentary and geochemical features of the clays indicate that they were laid down in a shallow body of water that alternated between

fresh and intensely saline-alkaline, probably as a response to changes in outflow geometry controlled by local volcanism¹⁵. Additional intermittent sources of fresh water are suggested by localized ephemeral stream channel features and the remains of an extensive fossil spring. Loc. 99 has produced fragmentary fossils representing suids, bovids, rodents, cercopithecoid primates and catfish. Eight additional faunal collecting areas in K3' have also produced elephants, hippopotami, carnivores, crocodiles, turtles, gastropods and additional micromammals. Many K3' taxa, notably hippopotami (*Hippopotamus*), crocodiles, catfish (*Clarias*), gastropods and turtles, reflect local aquatic conditions. The bulk of K3' non-aquatic fauna, including a colobine monkey, the elephant, the bovids *Kobus*, *Tragelaphus* and specimens probably belonging to *Syncerus*, and the suids *Potamochoerus porcus* (bushpig) and the extinct *Kolpochoerus majus*¹⁶, are consistent with a closed environment. The presence of the cane rat (*Thryonomys*) indicates localized patches of moist, marshy conditions.

Remains of *Homo* (KNM-BK 63-67 and KNM-BK 8518) were recovered at sites GnJh-01 and GnJh-19 by previous workers^{11,13,14} from K3 fluvial sediments to the west that represent a system of braided streams, some of which seem to have debouched into the lake. Fluvial K3 deposits and lacustrine K3' deposits are interstratified, indicating a shoreline that shifted in position in response to alterations in lake levels. The similarity in the array of fossils encountered in K3 and K3' sediments suggests that Middle Pleistocene *Pan* and *Homo* lived, or at least died, in broadly similar environmental settings. Taken together, the evidence suggests a locally wooded habitat on the shore of an alternately fresh and saline-alkaline lake, fluctuating lake levels, ephemeral nearshore fluvial channels, a nearby freshwater spring, and a semi-arid climatic regime. These conditions are not unlike those found near the shore of Lake Baringo today, although dense human populations have eliminated much of the woodland that formerly supported chimpanzees and the faunal community of which they were a part.

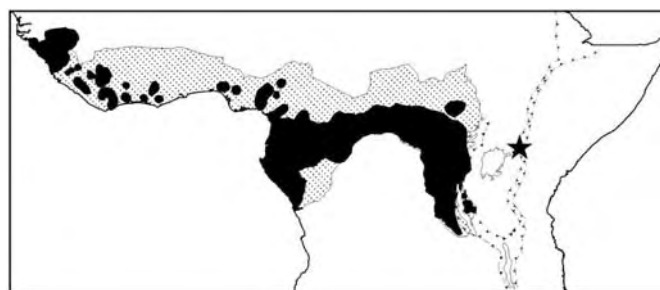


Figure 1 | Map showing current (solid black) and historical (stippled) ranges of *Pan* in equatorial Africa relative to major features of the eastern and western Rift Valleys. The Kapthurin Formation, Kenya, in the Eastern Rift Valley is marked by a star.

¹Department of Anthropology, University of Connecticut, Box U-2176, Storrs, Connecticut 06269, USA. ²Department of Anthropology, California Academy of Sciences, 875 Howard Street, San Francisco, California 94103, USA.

The chimpanzee specimens comprise a minimum of three teeth, probably from the same individual. Two of these are right and left upper central permanent incisors (I¹; KNM-TH 45519 and KNM-TH 45521, respectively). They exhibit broad, spatulate and moderately worn crowns, with thin dental enamel (Fig. 3). The lingual tubercle is large and flanked at the base by deep mesial and distal foveae, characteristic of *Pan*. This feature imparts great thickness to the labiolingual profiles of the teeth, and clearly distinguishes them from known hominins. The mesial and distal marginal ridges are well formed. The distal corners of the incisal edges are slightly chipped and the labial enamel surfaces exhibit pre-mortem wear as well as slight post-mortem surface weathering. The roots have closed apices and are straight, conical and relatively short. The incisal edges and lingual tubercles exhibit dentinal exposure resulting from wear. Measurements of the specimens, with comparisons to those of extant species of *Pan*, are provided in Table 1. The upper incisors are nearly identical to those of modern *Pan* in all aspects of morphology except their shorter root length. The sub-parallel mesial and distal margins of the incisors bestow a quadrate, rather than triangular, outline to the crowns, a feature that among living chimpanzees is considered to be more common among living *P. troglodytes* than *P. paniscus*¹⁷. The enamel and cementum coverings are in good condition and the perikymata on the labial surfaces of the crowns and the periradicular striae on the lingual surfaces of the roots can be easily seen. Several of the perikymata near the cervices of the teeth are faintly incised, indicating mild enamel hypoplasia having occurred at about the age of 5 years¹⁸. The well-matched mesial interproximal wear facets of the Kapthurin Formation *Pan* incisors (KNM-TH 45519 and KNM-TH 45521), the comparable degree of wear on their incisal edges, and the continuity of the enamel hypoplasia on their crowns and the incremental markings on their roots suggest that the two teeth are antimeres.

The third tooth is a lightly worn crown of a left upper permanent molar (KNM-TH 45520) (Fig. 4). It can be problematic to dis-

tinguish first from second upper molars in *Pan*, but we identify KNM-TH 45520 as an M¹, judging from the relatively large size of its hypocone, as this cusp is known to decrease in size from M¹ to M³ (ref. 19). The Kapthurin Formation M¹ is an extremely low molar crown that has lost most of the enamel on its mesial and lingual faces due to breakage after fossilization. The enamel surfaces are pock-marked as a result of chemical and physical weathering. The paracone and metacone are of approximately equal heights and are separated by a sharply incised buccal groove. The hypocone is lower than either of the buccal cusps, but is relatively large and well defined. A shallow trigon basin is delimited by a weak and obliquely oriented post-protocrista (crista obliqua). A deep but short distal fovea lies between the postprotocrista and the low distal marginal ridge. Despite marring of the enamel surface, perikymata are visible on the buccal and distal faces of the paracone, but there is no evidence of enamel hypoplasia. The relative thinness of the enamel can be discerned on

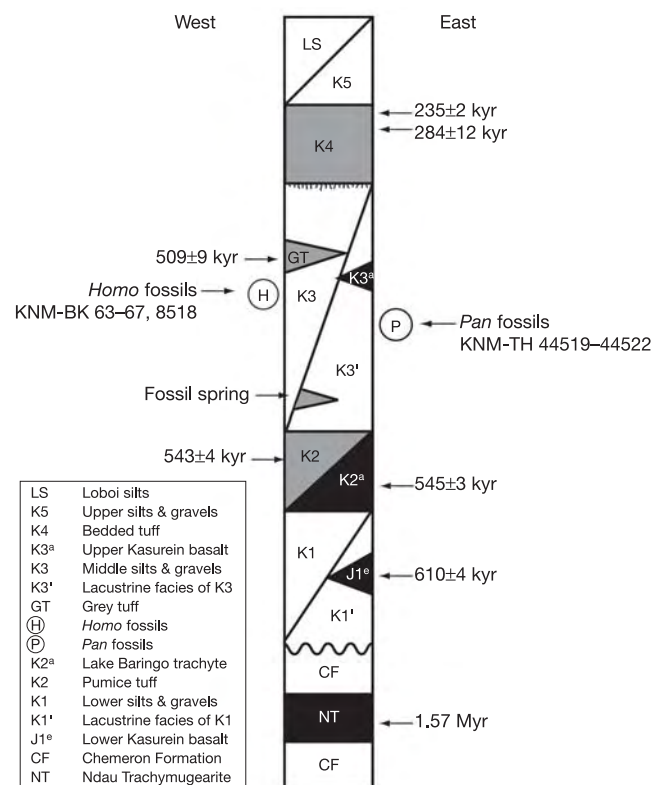


Figure 2 | Idealized stratigraphic column of the Kapthurin Formation, Kenya.



Figure 3 | Central upper incisors of *Pan* from the Kapthurin Formation, Kenya. **a**, KNM-TH 45519. From left to right: labial, lingual, mesial, distal and incisal views. **b**, KNM-TH 45521. Images are in the same sequence as for the previous specimen. **c**, Enlargement of the incisal edge of KNM-TH 45519 (left) and KNM-TH 45521 (right), showing the extreme thinness of the enamel characteristic of modern chimpanzees. **d**, Labial and lingual views of KNM-TH 45519 and KNM-TH 45521.

Table 1 | Dimensions of the Kapthurin Formation fossil chimpanzee teeth

Sample	Tooth	Mesiodistal dimension (mm)	Mesiodistal range (mm)	Buccolingual dimension (mm)	Buccolingual range (mm)
KNM-TH 45519	Right I ¹	10.46	—	9.12	—
KNM-TH 45521	Left I ¹	10.50	—	9.33	—
<i>P. troglodytes</i> (male)	I ¹	12.6 (n = 14)	10.5–13.5	10.1 (n = 15)	9.0–11.3
<i>P. troglodytes</i> (female)	I ¹	11.9 (n = 51)	10.0–13.4	9.6 (n = 50)	8.3–11.7
<i>P. paniscus</i> (male)	I ¹	10.3 (n = 15)	8.9–11.9	7.9 (n = 15)	7.2–9.2
<i>P. paniscus</i> (female)	I ¹	10.4 (n = 20)	9.0–11.5	7.6 (n = 21)	6.8–8.5
KNM-TH 45520	Left M ¹	9.7 (estimate)	—	Damage prevents measurement	—
<i>P. troglodytes</i> (male)	M ¹	10.3 (n = 19)	9.3–11.2	11.7 (n = 19)	10.7–13.2
<i>P. troglodytes</i> (female)	M ¹	10.1 (n = 51)	9.0–11.9	10.9 (n = 50)	7.0–12.8
<i>P. paniscus</i> (male)	M ¹	8.5 (n = 7)	7.9–9.4	9.5 (n = 6)	9.2–10.4
<i>P. paniscus</i> (female)	M ¹	8.3 (n = 6)	7.6–8.8	9.7 (n = 6)	9.3–10.4

Comparative dimensions are given for modern *P. troglodytes* and *P. paniscus* from ref. 19.

the broken mesial and lingual faces of the tooth. The extremely low height of the M¹ crown and the pronounced thinness of the enamel distinguish the tooth from those of known fossil or modern hominins. Among living chimpanzees, the presence of a well-expressed hypocone is more common in *P. troglodytes* than in *P. paniscus*²⁰. A fourth tooth (KNM-TH 45522), the crown and proximal roots of a tooth that may be plausibly identified as an aberrant right upper third molar (M³), will be described elsewhere and is not further discussed here.

The state of wear on the incisors and the M¹ conforms to the known sequence of dental emergence in *Pan*^{19,21}, and it is likely that they come from the same individual. If they do represent the same animal, its age at death can be estimated at approximately 7–8 years based on standards derived from captive animals²² and known dental maturation schedules for mandibular molars²³. The presence of linear enamel hypoplasia on the incisors, but not on the molars, is common in modern apes and seems to be related to nutritional stress that is experienced by the animal after weaning²⁴.

The morphology of the Kapthurin Formation teeth, especially the pronounced lingual tubercle on the incisors, the thickness of the bases of the incisors, the lowness of the molar crown, and the thinness of the enamel on all the teeth clearly supports their attribution to *Pan* rather than *Homo*. Specific diagnosis of isolated teeth within *Pan*, however, must be approached with caution, and for this reason we assign the Kapthurin Formation specimens to *Pan* sp. indet. Non-metric characters that have been suggested as diagnostic criteria for *P. troglodytes*, such as a more quadrilateral outline shape to the upper central incisor crowns¹⁷ and a better expressed hypocone on the maxillary molars^{19,20}, seem to suggest more similarity for the Kapthurin Formation fossils to *P. troglodytes* than to *P. paniscus*, but these features are variably expressed among the living species and

subspecies of *Pan*^{19,25}. Although mean tooth size is known to be significantly smaller in *P. paniscus* than in *P. troglodytes*^{17,25,26}, size ranges overlap (Table 1). Furthermore, apart from the present specimens, we lack a fossil record for the Pliocene and Pleistocene from which to assess past variability within the genus, and it is feasible that the Kapthurin Formation fossils represent members of an extinct lineage within the genus *Pan*.

The Kapthurin Formation fossils represent the first unequivocal evidence of *Pan* in the fossil record, and they demonstrate the presence of chimpanzees in the eastern Rift Valley of Kenya, ~600 km east of the limit of their current range (Fig. 1). The Rift Valley clearly did not pose a physiographical or ecological barrier to chimpanzee occupation. Chimpanzee habitat is now highly fragmented, in part by human activities, but in historic times chimpanzees ranged over a wide belt of equatorial Africa from southern Senegal to western Uganda and Tanzania (Fig. 1). Although much of this region is rainforest, chimpanzees currently also occupy dry forest, woodland and dry savannah, particularly near the eastern edge of their range^{27–29}. The modern Baringo region ecosystem is a mosaic of semi-arid *Acacia* bushland and riverine woodland, with a significant substratum of perennial and annual grasses³⁰. The Tugen Hills palaeosol carbon isotope record indicates that the woodland and grassland components of the vegetation have been present there from 16 Myr³⁰. Representatives of both *Homo* and *Pan* are present in the same stratigraphic interval of the Kapthurin Formation at sites only ~1 km apart, and faunal data suggest that they occupied broadly similar environments in the Middle Pleistocene. This evidence shows that in the past chimpanzees occupied regions in which the only hominoid inhabitants were thought to have been members of the human lineage. Now that chimpanzees are known to form a component of the Middle Pleistocene fauna in the Rift Valley, it is quite

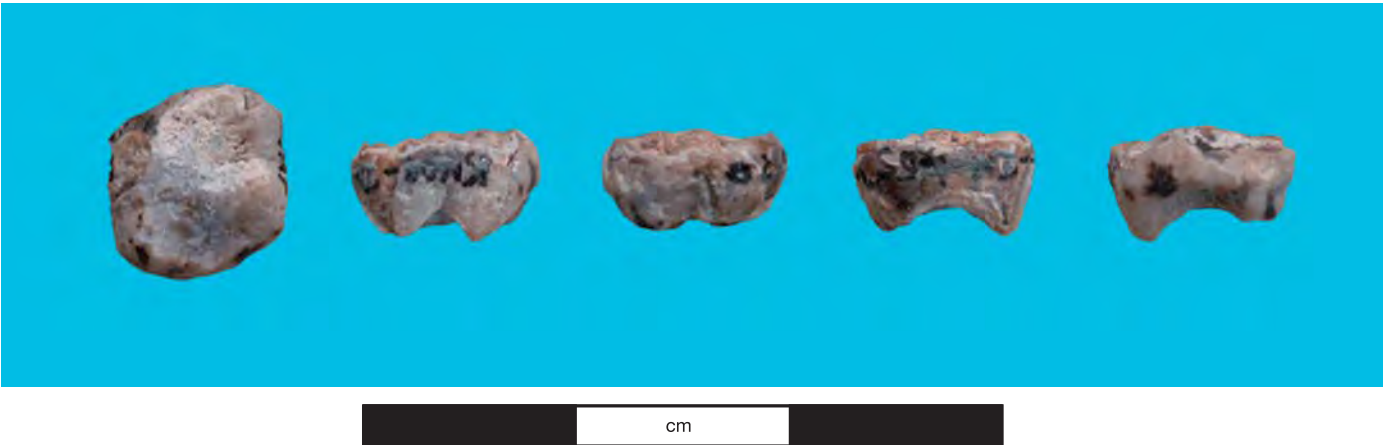


Figure 4 | Upper left first molar (KNM-TH 45520). From left to right: occlusal, labial, lingual, mesial and distal views. Note the thinness of the enamel on the broken mesial face of the paracone in the mesial view.

possible that they remain to be recognized in other portions of the fossil record there, and that chimpanzees and hominins have been sympatric since the time of their divergence.

Received 31 January; accepted 4 July 2005.

- Ruvolo, M. E. Molecular phylogeny of the hominoids: inferences from multiple independent DNA sequence data sets. *Mol. Biol. Evol.* **14**, 248–265 (1997).
- Darwin, C. *The Descent of Man and Selection in Relation to Sex* (John Murray, London, 1871).
- Washburn, S. L. in *Changing Perspectives on Man* (ed. Rothblatt, B.) 193–201 (Univ. Chicago Press, Chicago, 1968).
- Kortlandt, A. *New Perspectives on Ape and Human Evolution* (Univ. Amsterdam, Amsterdam, 1972).
- Pilbeam, D. & Young, N. Hominoid evolution: synthesizing disparate data. *C. R. Palevol.* **3**, 305–321 (2004).
- Coppens, Y. East side story: the origin of mankind. *Sci. Am.* **270**, 88–95 (1994).
- Martyn, J. *The Geologic History of the Country Between Lake Baringo and the Kerio River, Baringo District, Kenya* (PhD dissertation, Univ. London, 1969).
- Tallon, P. in *Geological Background to Fossil Man* (ed. Bishop, W. W.) 361–373 (Scottish Academic Press, Edinburgh, 1978).
- McBrearty, S., Bishop, L. C. & Kingston, J. Variability in traces of Middle Pleistocene hominid behaviour in the Kapthurin Formation, Baringo, Kenya. *J. Hum. Evol.* **30**, 563–580 (1996).
- McBrearty, S. in *Late Cenozoic Environments and Hominid Evolution: a Tribute to Bill Bishop* (eds Andrews, P. & Banham, P.) 143–156 (Geological Society, London, 1999).
- McBrearty, S. & Brooks, A. The revolution that wasn't: a new interpretation of the origin of modern human behaviour. *J. Hum. Evol.* **39**, 453–563 (2000).
- Deino, A. & McBrearty, S. $^{40}\text{Ar}/^{39}\text{Ar}$ chronology for the Kapthurin Formation, Baringo, Kenya. *J. Hum. Evol.* **42**, 185–210 (2002).
- Leakey, M., Tobias, P. V., Martyn, J. E. & Leakey, R. E. F. An Acheulian industry with prepared core technique and the discovery of a contemporary hominid at Lake Baringo, Kenya. *Proc. Prehist. Soc.* **35**, 48–76 (1969).
- Wood, B. A. & Van Noten, F. L. Preliminary observations on the BK 8518 mandible from Baringo, Kenya. *Am. J. Phys. Anthropol.* **69**, 117–127 (1986).
- Renaut, R. W., Tiercelin, J.-J. & Owen, B. in *Lake Basins Through Space and Time* (eds Gierlowski-Kordesch, E. H. & Kelts, K. R.) 561–568 (Am. Assoc. Petrol. Geol., Tulsa, Oklahoma, 2000).
- Bishop, L. C., Hill, A. P. & Kingston, J. in *Late Cenozoic Environments and Hominid Evolution: a Tribute to Bill Bishop* (eds Andrews, P. & Banham, P.) 99–112 (Geological Society, London, 1999).
- Johanson, D. C. Some metric aspects of the permanent and deciduous dentition of the pygmy chimpanzee (*Pan paniscus*). *Am. J. Phys. Anthropol.* **41**, 39–48 (1974).
- Dean, M. C. & Reid, D. J. Perikymata spacing and distribution on hominid anterior teeth. *Am. J. Phys. Anthropol.* **116**, 209–215 (2001).
- Swindler, D. R. *Primate Dentition: An Introduction to the Teeth of Non-Human Primates* (CUP, Cambridge, 2002).
- Kinzey, W. G. in *The Pygmy Chimpanzee* (ed. Susman, R. L.) 65–88 (Plenum, New York, 1984).
- Smith, B. H., Crummett, T. L. & Brandt, K. L. Ages of eruption of primate teeth: a compendium for aging individuals and comparing life histories. *Yearb. Phys. Anthropol.* **37**, 177–232 (1994).
- Kuykendall, K. L., Mahoney, C. J. & Conroy, G. C. Probit and survival analysis of tooth emergence ages in a mixed-longitudinal sample of chimpanzees (*Pan troglodytes*). *Am. J. Phys. Anthropol.* **89**, 379–399 (1992).
- Anemone, R. L., Watts, E. S. & Swindler, D. R. Dental development of known-age chimpanzees, *Pan troglodytes* (Primates, Pongidae). *Am. J. Phys. Anthropol.* **86**, 229–241 (1991).
- Skinner, M. F. & Hopwood, D. Hypothesis for the causes and periodicity of repetitive linear enamel hypoplasia in large, wild African (*Pan troglodytes* and *Gorilla gorilla*) and Asian (*Pongo pygmaeus*) apes. *Am. J. Phys. Anthropol.* **123**, 216–235 (2004).
- Uchida, A. *Craniodental Variation Among the Great Apes* (Harvard Univ. Peabody Mus., Cambridge, Massachusetts, 1996).
- Johanson, D. C. *An Odontological Study of the Chimpanzee with Some Implications for Hominoid Evolution* (PhD dissertation, Univ. Chicago, 1974).
- Kormos, R., Boesch, C., Bakarr, M. I. & Butynski, T. M. *West African Chimpanzees: Status Survey and Conservation Action Plan* (IUCN Publication Unit, Cambridge, 2003).
- McGrew, W. C., Baldwin, P. J. & Tutin, C. E. G. Chimpanzees in a hot, dry and open habitat: Mt. Assirik, Senegal. *J. Hum. Evol.* **10**, 227–244 (1981).
- McGrew, W. C., Marchant, L. F. & Nishida, T. *Great Ape Societies* (CUP, Cambridge, 1996).
- Kingston, J. D., Marino, B. & Hill, A. P. Isotopic evidence for Neogene hominid palaeoenvironments in the Kenya Rift Valley. *Science* **264**, 955–959 (1994).

Acknowledgements We wish to thank B. Kimeu, N. Kanyenze and M. Macharwas, who found the chimpanzee fossils reported here. Research in the Kapthurin Formation is carried out with the support of an NSF grant to S.M., and under a research permit from the Government of the Republic of Kenya and a permit to excavate from the Minister for Home Affairs and National Heritage of the Republic of Kenya. Both of these are issued to A. Hill and the Baringo Paleontological Research Project, an expedition conducted jointly with the National Museums of Kenya. We also thank personnel of the Departments of Palaeontology, Ornithology and Mammalogy of the National Museums of Kenya, Nairobi; A. Zihlman; and Y. Hailie-Selassie, L. Jellema and M. Ryan for curation and access to specimens. We express gratitude to A. Hill for his comments on the manuscript. We also thank G. Chaplin for drafting Fig. 1, B. Warren for preparing Figs 3 and 4, and A. Bothell for help with submission of the figures. We are grateful to J. Kelley, J. Kingston, M. Leakey, R. Leakey, C. Tryon, A. Walker and S. Ward for discussions. We thank G. Suwa for his remarks.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.M. (mcbrearty@uconn.edu).

A disk of dust and molecular gas around a high-mass protostar

Nimesh A. Patel¹, Salvador Curiel^{1,2}, T. K. Sridharan¹, Qizhou Zhang¹, Todd R. Hunter¹, Paul T. P. Ho^{1,3}, José M. Torrelles^{4,†}, James M. Moran¹, José F. Gómez^{5,6} & Guillem Anglada⁶

The processes leading to the birth of low-mass stars such as our Sun have been well studied¹, but the formation of high-mass (over eight times the Sun's mass, $M_{\odot}$) stars remains poorly understood². Recent studies suggest that high-mass stars may form through accretion of material from a circumstellar disk³, in essentially the same way as low-mass stars form, rather than through the merging of several low-mass stars⁴. There is as yet, however, no conclusive evidence^{5,6}. Here we report the presence of a flattened disk-like structure around a massive $15M_{\odot}$ protostar in the Cepheus A region, based on observations of continuum emission from the dust and line emission from the molecular gas. The disk has a radius of about 330 astronomical units (AU) and a mass of 1 to 8 $M_{\odot}$. It is oriented perpendicular to, and spatially coincident with, the central embedded powerful bipolar radio jet, just as is the case with low-mass stars, from which we conclude that high-mass stars can form through accretion.

Previously reported disk-like structures associated with high-mass stars, which are more than twice as far away as low-mass stars, have typical sizes of several thousands of AU. Furthermore, in such sources, the presence of thermal jets at scales of a few hundred AU has not been demonstrated, in part owing to the lack of sufficient angular resolution. Consequently, it has not been possible to identify and isolate a good example of a high-mass protostar with a disk-jet system^{7–9}. The recently reported massive disk in M17-S01 (ref. 5), based on near-infrared observations of the disk in silhouette against the bright background light of the ionized region and in emission lines of carbon monoxide (CO), has subsequently been shown to be in fact a much lower-mass disk ($0.09M_{\odot}$; ref. 6). In the former study⁵, uncertainty resulted from (1) insufficient angular resolution ($8''$), (2) observations of molecular species that are probably optically thick and tracers of low-density gas (CO and its isotopes, ^{13}CO and C^{18}O), and (3) the absence of an independent means of inferring the presence of a high-mass protostar, such as bright radio continuum emission and/or luminous water maser sources.

Cepheus A is a well-studied high-mass star-forming region in the Cepheus OB3 complex^{10,11}. The bolometric luminosity of this high-mass star-forming region is about $2.5 \times 10^4 L_{\odot}$ (ref. 12). Half of this luminosity is attributed to the HW2 object, the brightest radio continuum source in the field, which is considered to be a B0.5 spectral type protostar of $15M_{\odot}$ (refs 13–16) and is the most probable exciting source of a powerful extended bipolar molecular outflow^{17,18}. The radio continuum flux density of about 40 mJy at 1.3-cm wavelength¹⁶ would correspond to a flux density of about 0.8 Jy if the source were at the distance of a typical low-mass star-forming region such as the Taurus molecular cloud (160 pc). This would be

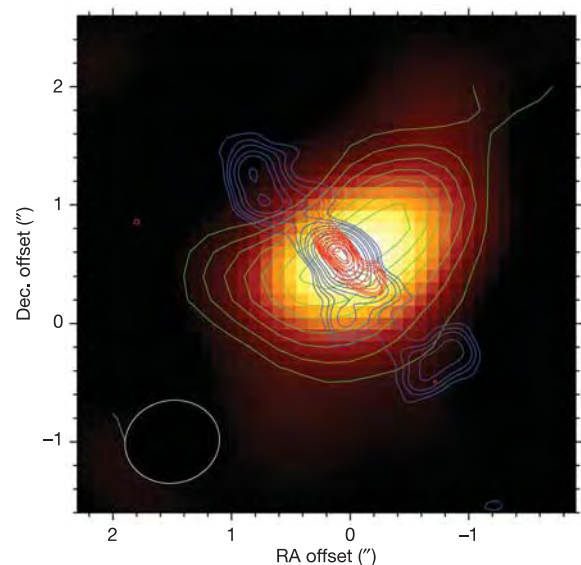


Figure 1 | Emission from the Cepheus A HW2 protostar. Dust continuum emission at 327 GHz is shown in the image, ranging linearly from 0 to 1.5 Jy beam^{−1}, and the integrated intensity in the $\text{CH}_3\text{CN } J = 18-17$ ($K = 0, 1, 2, 3$) line emission from -35 to 30 km s^{-1} line-of-sight velocities is shown with the contour (levels from 5 to 40 Jy beam^{−1} km s^{−1} in steps of 5 Jy beam^{−1} km s^{−1}). CH_3CN has been shown to trace high-density (10^6 cm^{-3}) gas in massive star-forming regions^{25,30}. The SMA beam size was $0.8'' \times 0.7''$ with a position angle of -78.6° (left lower corner). This angular resolution is by far the highest reached at submillimetre wavelength observations (for example, compared to single-dish observations, we have more than an order of magnitude greater angular resolution). Blue contours show the 3.6-cm-wavelength continuum emission from a well-collimated jet, whereas red contours show the 1.3-cm-wavelength continuum emission from the inner part of the jet^{16,20}. The protostar is believed to be located at the origin of this elongated jet. The absolute astrometric error in the alignment of the VLA and SMA observations is $\approx 0.2''$. The elongation in both dust and CH_3CN emission is nearly perpendicular to the circumstellar thermal jet, strongly supporting the disk interpretation. The deconvolved disk radius is $\approx 330 \text{ AU}$. The submillimetre observations were carried out using seven of the eight available antennas of the SMA in the extended array configuration with a maximum baseline of 226 m. The observations were made on 30 August 2004. With the receivers tuned to 321 GHz, we had the $\text{CH}_3\text{CN } J = 18-17$, $K = 0, 1, 2, 3, \dots, 9$ lines in the upper sideband and the $10_{29-9_{36}}$ water maser transition in the lower sideband. Submillimetre water masers were found to be associated with the HW2 and HW3c sources (N.A.P. *et al.*, manuscript in preparation).

¹Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, MS78, Cambridge, Massachusetts 02138, USA. ²Instituto de Astronomía, UNAM, Apartado Postal 70-264, 04510 Mexico DF, Mexico. ³Academia Sinica Institute of Astronomy and Astrophysics, Taipei, Taiwan. ⁴Consejo Superior de Investigaciones Científicas-IEEC, Gran Capita 2-4, E-08034 Barcelona, Spain. ⁵Laboratorio de Astrofísica Espacial y Física Fundamental, INTA, Apartado 50727, E-28080 Madrid, Spain. ⁶Instituto de Astrofísica de Andalucía, CSIC, Apartado 3004, Camino Bajo de Huetor 50, E-18008 Granada, Spain. [†]Present address: UK Astronomy Technology Centre, Royal Observatory Edinburgh, Blackford Hill, Edinburgh EH9 3HJ, UK (J. M. T.).

about a hundred times brighter than typical low-mass protostars at that distance. Also associated with HW2 is a 'biconical thermal radio jet' at the size scale of $\approx 1'' (\approx 725 \text{ AU})$, with the ionized gas exhibiting proper motions with velocities $\geq 500 \text{ km s}^{-1}$ along the axis of the jet^{14,19,20}. In addition, we estimate that the luminosity of the masers associated with HW2 (ref. 21) is $L_{\text{H}_2\text{O}} \geq 3 \times 10^{-6} L_{\odot}$, which is $\sim 10^3$ to 10^5 times more luminous than water masers associated with low-mass stars²². These comparisons strongly imply that HW2 is a high-mass protostar.

Our new sub-arcsecond angular resolution Submillimeter Array (SMA)²³ observations towards HW2 have spatial resolution and sensitivity to dust emission far superior to those of existing sub-millimetre observations made with single-dish telescopes or interferometric observations made at millimetre and centimetre wavelengths. This allowed us to image directly the compact circumstellar disk surrounding HW2 at the scale of a few hundred AU, and to define its physical properties. Figure 1 shows an overlay of the dust continuum emission at 327 GHz and the integrated intensity in the methyl cyanide $\text{CH}_3\text{CN } J=18-17$ line emission. The dust continuum emission was imaged by selecting spectrometer channels free of line emission. The deconvolved source radius is 330 AU ($0.45''$) in dust continuum emission, and 580 AU ($0.8''$) in CH_3CN line emission. The integrated flux density from the dust emission is $2.0 \text{ Jy} \pm 0.1 \text{ Jy}$. The gaussian-fitted positions of the dust and CH_3CN emission peaks agree well to within a tenth of an arcsecond. The sizes, orientations and other characteristics are summarized in Table 1. By examining the visibility amplitudes as a function of baseline distance, we have confirmed that this flattened structure is well resolved along its major axis and partially resolved along its minor axis.

The free-free continuum emission from the thermal jet at 1.3- and 3.6-cm wavelengths, observed with the Very Large Array (VLA), is also shown in Fig. 1. The 1.3-cm wavelength continuum emission traces the inner part of the jet, with the protostellar source most probably at its geometric centre^{16,19}. The position angle of the thermal jet at both 1.3- and 3.6-cm wavelengths is $\sim 45^\circ$. A large-scale extended ($\sim 1'$) bipolar outflow mapped in $\text{HCO}^+ J=1-0$ emission is centred at the position of HW2 (ref. 17) and has the same orientation as the much smaller-scale centimetre wavelength continuum jet. The size and morphology of the dust and gas emission are in good agreement with each other and the elongation in both these emissions is seen to be nearly perpendicular to, and peaking on, the biconical circumstellar thermal jet. This strongly supports the disk interpretation for the flattened structure seen in the dust and CH_3CN emission. From these observations of the jet and outflow at various wavelengths, the existence of such a disk might be expected, since outflows in low-mass stars are launched from the inner portions of the rotating disks according to theoretical models (see ref. 24). Although these models are for low-mass stars, our observational results suggest that outflows in high-mass stars may be produced in a similar fashion.

Figure 2 shows a position-velocity diagram along the major axis of the elongated emission seen in the CH_3CN lines. From this diagram we see a velocity gradient of $\approx 6 \text{ km s}^{-1}$ over $0.5''$ (considering the largest velocity shift from the central position). We interpret this

velocity gradient to be due to gravitationally bound rotational motion. The dynamical mass enclosed within a radius r with rotational velocity v_r is given by $v_r^2 r / G$, where G is the gravitational constant. The observed velocity (along the line of sight) is $v = v_r \sin(i)$, where i is the inclination angle of the disk axis with respect to the line of sight. From the mean value of the observed aspect ratio of gas and dust disk as listed in Table 1 (that is, assuming a circular disk) we estimate the inclination angle to be $\approx 62^\circ$, and therefore a binding mass of $19 \pm 5 M_{\odot}$. This implies that the observed motions can be bound by the central high-mass protostar. In addition, from the line ratios of the $\text{CH}_3\text{CN } K$ components we can estimate the temperature of the gas²⁵. We assume optically thin emission and a single value of temperature along the line of sight. From the ratios of the peak intensities of the $K=3$ and $K=2$ components, we find that the temperature varies along the major axis of the disk from ~ 25 to $\sim 75 \text{ K}$, with the northwest part of the disk relatively cooler and the southeast and central parts of the disk relatively hotter. The linewidth also appears to be greater at the latter positions.

With the present angular resolution we are unable to map in detail the temperature and kinematics of the gas in the disk, but we can make rough estimates of its mass. If we assume that the dust is optically thin, we can estimate the gas mass of the disk²⁶ assuming the gas-to-dust ratio of 100, gas temperature of 50 K (from analysis of methyl cyanide lines) and the measured 900 μm continuum flux density of 2.0 Jy. We also need to know the grain emissivity spectral index β for the dust emission. We obtain a mass estimate for the gas disk to be $1 M_{\odot}$ for $\beta = 1$ (ref. 27) and $8 M_{\odot}$ for $\beta = 2$ (ref. 28). Some of this mass could be in a rotationally flattened infalling envelope instead of being in the disk but we note that the observed size of the disk in HW2 is in good agreement with the range of ≈ 400 to 600 AU for the centrifugal radius (the radius where disk formation is expected to occur) derived from a recent fitting of flattened infalling envelope models to the observed spectral energy distribution of high-

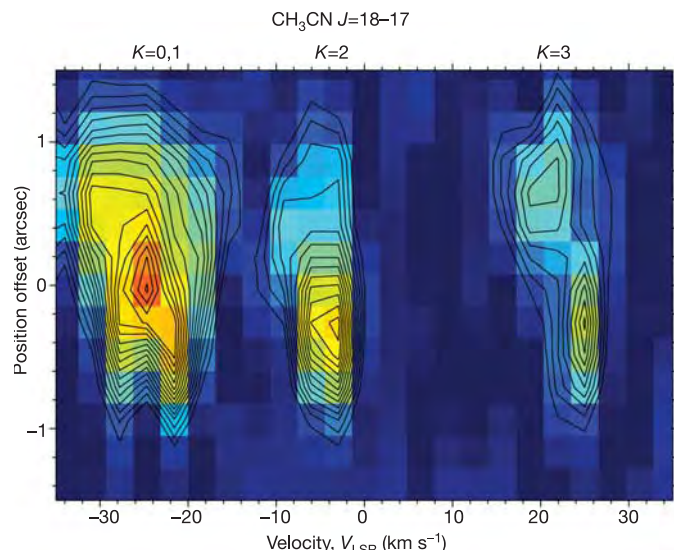


Figure 2 | Position-velocity map of CH_3CN emission along the major axis of the elongated structure shown in Fig. 1. The position angle is -53° (with respect to North, with East as positive). The contour levels are from 0.3 to 1.5 Jy beam^{-1} every 0.1 Jy beam^{-1} . The position offset is measured along the major axis, with positive offset corresponding to the southeast part and negative offset towards the northwest part of the disk. The $0''$ position offset corresponds to the peak of the integrated intensity. The $K=0$ and 1 components are blended. $K=4$ and higher lines were not detected. The reference frequency was chosen to be at the $J=18-17$, $K=2$ line of CH_3CN . We interpret the velocity shift of $\sim 6 \text{ km s}^{-1}$ over $0.5''$ seen in the $K=2$ and $K=3$ emission to be due to rotational motion and estimate a binding mass of $19 \pm 5 M_{\odot}$. LSR, local standard of rest.

Table 1 | Characteristics of Cepheus A HW2 dust continuum and CH_3CN emission

	Total flux density	Major axis ($''$)	Minor axis ($''$)	PA ($^\circ$)	$\Delta\alpha(^{\circ})$	$\Delta\delta(^{\circ})$
Continuum	$2.0 \pm 0.1 \text{ Jy}$	0.9	0.5	-59.2 ± 0.6	-0.07	0.55
CH_3CN	122 Jy km s^{-1}	1.6	0.6	-56.2 ± 0.2	-0.04	0.58

Deconvolved gaussian-fitted model parameters for the Cepheus A HW2 disk structure seen in continuum and CH_3CN emission at 330 GHz ($J=18-17$, $K=0, 1, 2, 3$). The last two columns are the positions of the peaks with respect to $\alpha(2000) = 22 \text{ h } 56 \text{ min } 17.970 \text{ s}$, $\delta(2000) = +62^\circ 01' 48.992''$. The uncertainties in the sizes and positions are $\sim 0.1''$.

mass protostars²⁹. All these results support theoretical models of high-mass star formation via an accretion process occurring in a disk around the protostar, accompanied by a bipolar outflow (much like low-mass stars), rather than by models that require merging of several low-mass stars.

Received 25 May; accepted 28 June 2005.

1. Lada, C. J. & Shu, F. H. The formation of sunlike stars. *Science* **248**, 564–572 (1990).
2. Stahler, S. W., Palla, F. & Ho, P. T. P. in *Protostars and Planets IV* (eds Mannings, V., Boss, A. P. & Russell, S.) 327–351 (Univ. Arizona Press, Tucson, 2000).
3. McKee, C. & Tan, J. The formation of massive stars from turbulent cores. *Astrophys. J.* **585**, 850–871 (2003).
4. Bonnell, I. A. & Bate, M. R. Accretion in stellar clusters and the collisional formation of massive stars. *Mon. Not. R. Astron. Soc.* **336**, 659–669 (2002).
5. Chini, R. *et al.* The formation of a massive protostar through the disk accretion of gas. *Nature* **429**, 155–157 (2004).
6. Sako, S. *et al.* No high-mass protostars in the silhouette young stellar object M17-SO1. *Nature* **434**, 995–998 (2004).
7. Zhang, Q., Hunter, T. R. & Sridharan, T. K. A rotating disk around a high-mass young star. *Astrophys. J.* **505**, L151–L154 (1998).
8. Shepherd, D., Claussen, M. J. & Kurtz, S. E. Evidence for a solar system-size accretion disk Around the massive protostar G192.16–3.82. *Science* **292**, 1513–1518 (2001).
9. Cesaroni, R. *et al.* Unveiling the disk-jet system in the massive (proto)star IRAS 20126+4104. *Astron. Astrophys.* **345**, 949–964 (1999).
10. Sargent, A. I. Molecular clouds and star formation. I — Observations of the Cepheus OB3 molecular cloud. *Astrophys. J.* **218**, 736–748 (1977).
11. Blitz, L. & Lada, C. J. H₂O masers near OB associations. *Astrophys. J.* **227**, 152–158 (1979).
12. Evans, N. J. *et al.* Far-infrared observations of the Cepheus OB3 molecular cloud. *Astrophys. J.* **244**, 115–123 (1981).
13. Hughes, V. A. & Wouterloot, J. G. A. The star-forming region in Cepheus A. *Astrophys. J.* **276**, 204–210 (1984).
14. Rodríguez, L. F. *et al.* Cepheus A HW2: A powerful thermal radio jet. *Astrophys. J.* **430**, L65–L68 (1994).
15. Hughes, V., Cohen, R. & Garrington, S. High-resolution observations of Cepheus A. *Mon. Not. R. Astron. Soc.* **272**, 469–480 (1995).
16. Torrelles, J. M. *et al.* The thermal radio jet of Cepheus A HW2 and the water maser distribution at 0.0.8" scale (60 AU). *Astrophys. J.* **L107–L111** (1996).
17. Gómez, J. F. *et al.* Disk and outflow in Cepheus A-HW2: Interferometric SiO and HCO⁺ observations. *Astrophys. J.* **514**, 287–295 (1999).
18. Rodríguez, L. F., Ho, P. T. P. & Moran, J. Anisotropic mass outflow in Cepheus A. *Astrophys. J.* **240**, L149–L152 (1980).
19. Curiel, S. *et al.* Large proper motions in the jet of the high-mass YSO Cepheus A HW2. *Astrophys. J.* (submitted).
20. Rodríguez, L. F., Torrelles, J. M., Anglada, G. & Martí, J. VLA observations of brightness enhancements moving along the axis of the Cep A HW2 thermal jet. *Rev. Mex. Astron. Astrofis.* **37**, 95–99 (2001).
21. Torrelles, J. M. *et al.* Systems with H₂O maser and 1.3 centimeter continuum emission in Cepheus A. *Astrophys. J.* **509**, 262–269 (1998).
22. Furuya, R. S., Kitamura, Y., Wootten, A., Claussen, M. J. & Kawabe, R. Water maser survey toward low-mass young stellar objects in the northern sky with the Nobeyama 45 meter telescope and the very large array. *Astrophys. J. Suppl.* **144**, 71–134 (2003).
23. Ho, P. T. P., Moran, J. M. & Lo, K. Y. The Submillimeter Array. *Astrophys. J.* **616**, L1–L6 (2004).
24. Shu, F., Najita, J., Ostriker, E. & Shang, H. Magnetocentrifugally driven flows from young stars and disks. V. Asymptotic collimation into jets. *Astrophys. J.* **455**, L155–L158 (1996).
25. Loren, R. B. & Mundy, L. G. The methyl cyanide hot and warm cores in Orion — Statistical equilibrium excitation models of a symmetric-top molecule. *Astrophys. J.* **286**, 232–251 (1984).
26. Hildebrand, R. The determination of cloud masses and dust characteristics from submillimetre thermal emission. *Q. J. R. Astron. Soc.* **24**, 267–282 (1983).
27. Williams, S. J., Fuller, G. A. & Sridharan, T. K. The circumstellar environments of high-mass protostellar objects. I. Submillimetre continuum emission. *Astron. Astrophys.* **417**, 115–133 (2004).
28. Hunter, T. *et al.* 350 micron images of massive star formation regions. *Astron. J.* **119**, 2711–2727 (2000).
29. De Buizer, J., Osorio, M. & Calvet, N. Observations and modeling of the 2–25 μ m emission from high mass protostellar object candidates. *Astrophys. J.* (in the press).
30. Pankonin, V., Churchwell, E., Watson, C. & Bieging, J. H. A methyl cyanide search for the earliest stages of massive protostars. *Astrophys. J.* **558**, 194–203 (2001).

Acknowledgements S.C. acknowledges support from DEGAPA/UNAM and CONACYT grants and from the Submillimeter Array project. G.A., J.F.G. and J.M.T. are supported by a grant (including FEDER funds) of the Ministerio de Educación y Ciencia, Spain. G.A. acknowledges support from Junta de Andalucía. The Submillimeter Array is a joint project between the Smithsonian Astrophysical Observatory and the Academia Sinica Institute of Astronomy and Astrophysics, and is funded by the Smithsonian Institution and the Academic Sinica. We are grateful to the people of Hawai'ian ancestry on whose sacred mountain (Mauna Kea) we are privileged to be guests.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to N.A.P. (npatel@cfa.harvard.edu).

LETTERS

A circumstellar disk associated with a massive protostellar object

Zhibo Jiang¹, Motohide Tamura², Misato Fukagawa², Jim Hough³, Phil Lucas³, Hiroshi Suto², Miki Ishii² & Ji Yang¹

The formation process for stars with masses several times that of the Sun is still unclear. The two main theories are mergers of several low-mass young stellar objects¹, which requires a high stellar density, or mass accretion from circumstellar disks in the same way as low-mass stars are formed², accompanied by outflows during the process of gravitational infall. Although a number of disks have been discovered around low- and intermediate-mass young stellar objects^{3,4}, the presence of disks around massive young stellar objects is still uncertain and the mass of the disk system detected around one such object⁵, M17, is disputed⁶. Here we report near-infrared imaging polarimetry that reveals an outflow/disk system around the Becklin–Neugebauer protostellar object, which has a mass of at least seven solar masses ($M_{\odot}$). This strongly supports the theory that stars with masses of at least $7M_{\odot}$ form in the same way as lower mass stars.

The Becklin–Neugebauer object (BN; ref. 7) is a famous candidate protostar located within the central part of the Orion molecular cloud (OMC-1). With a distance of $\sim 1,500$ light-years from the Sun ($\sim 1.39 \times 10^{19}$ m; ref. 8), this object has a luminosity of $\sim 2,500$ solar luminosities ($L_{\odot}$), but may be up to $\sim 10^4 L_{\odot}$ taking its associated luminosity into account⁹. A few studies suggest a mass of $\sim 7M_{\odot}$ (refs 10, 11), but it may be as much as $20M_{\odot}$ (ref. 11). The object is red in the near-infrared with $[H - K] \approx 3.8$ mag ($\lambda_H = 1.65 \mu\text{m}$, $\lambda_K = 2.2 \mu\text{m}$). Even after correction for a line-of-sight visual extinction of $A_V \approx 17$ (ref. 9), it has a significantly large colour excess, suggesting the presence of circumstellar materials.

To further investigate the properties of BN, we report here H-band and K-band high-resolution (full-width at half-maximum, FWHM = $0.1''$) polarimetric imaging. Figure 1 shows the images of polarization degrees superposed by corresponding brightness contours and polarization vectors in the H- and K-bands in the vicinity of BN. A butterfly-shaped bright pattern and a flare-shaped dark lane running from the southeast to the northwest at a position angle of $\sim 126^\circ$ (PA, measured from north to east) are clearly seen in the H-band (Fig. 1a), but are less clear in the K-band. This structure can be interpreted as an outflow/disk system around the star.

There are two main processes for producing polarization: differential absorption by non-spherical dust grains with the short axis preferentially aligned along the local magnetic field, referred to as dichroic extinction¹², and by scattering. Dichroic absorption alone is unlikely to give the observed polarization morphology, because this would require a sandwich-like dust distribution with the optical depth being lower in the central lane than in the two wings^{13,14}, and the shape of the bright wings is quite regularly parabolic, which is unlikely to arise from extinction. The highly polarized bipolar feature in Fig. 1a is most easily interpreted as the cavity walls formed by a bipolar outflow from BN. Our unpublished polarimetric data of long

exposure but lower spatial resolution in the K-band indicate reflection nebulosities $\sim 5''$ away to the northeast and southwest, which are illuminated by BN. Other evidence for a bipolar outflow comes from the presence of local H_2 emission to the northeast of BN^{15,16}. The dark lane, which is nearly perpendicular to the bipolar outflow, almost certainly represents the circumstellar disk around the object. Supporting evidence can be found from the $12.5\text{-}\mu\text{m}$ image in ref. 17, in which the mid-infrared emission of BN is elongated in the same direction, although the disk is not resolved. In view of the roughly symmetric polarization morphology, the disk axis should be lying nearly in the plane of the sky. This kind of configuration has been observed around low-mass young stellar objects such as HL Tau¹⁸, and Herbig Ae/Be stars⁵.

To show that a disk/bipolar outflow system can create such a polarization structure, we use a Monte Carlo code for dust scattering to simulate the emerging photons from the system. The resultant polarization image from the scattering model is shown in Fig. 2a. The polarization image clearly shows a butterfly structure with a dark lane running between two wings. This structure is quite similar to the observed features in the H-band, thus confirming the presence of a circumstellar disk around the object.

Despite the overall similarity between the observations and the model, there are some obvious differences. Instead of a centrosymmetric pattern, the observed polarization vectors appear parallel. This is different to low-mass young stellar objects, where centrosymmetric polarization patterns dominate the reflection nebulae. Here, dichroic absorption may be producing significant polarization. Observations have shown that the foreground magnetic field is oriented at $PA \approx 120^\circ$ (refs 14, 15, 19), and non-spherical dust grains will produce absorptive polarization vectors parallel to the field. The overall effect will be to produce a pattern of essentially parallel polarization vectors, with degrees of polarization higher in the H-band than in the K-band. To explore this possibility, we use a virtual partial polarizer, with its main axis aligned parallel to the disk plane, in front of the scattering model. Assuming the polarizer can produce 20% polarized light, we successfully modified the centrosymmetric pattern, as shown in Fig. 2b. It is therefore possible to interpret the observed polarization in terms of the combined effect of dust scattering and dichroic extinction.

Another interesting feature is that the polarization degrees in the two wings are quite different. The polarization degrees in the northeast are larger than those in the southwest by $\sim 20\%$. Figure 3 shows the polarization degrees changing with the distance from the central object along the axis ($PA \approx 36^\circ$) and disk plane for the H-band. Monte Carlo simulations with different disk inclinations either change the polarization morphology significantly or do not yield such large differences in polarization degree. Therefore it is not likely that the difference comes from the viewing angle of the disk. The

¹Purple Mountain Observatory, Chinese Academic of Sciences, Nanjing 210008, China. ²National Astronomical Observatories of Japan, Osawa 2-21-1, Mitaka, Tokyo 181-8588, Japan. ³Department of Physics, Astronomy & Mathematics, University of Hertfordshire, College Lane, Hatfield AL10 9AB, UK.

most likely interpretation is differential extinction or different dust densities between the two wings.

The orientation of the system is interesting: the axis of the disk and the outflow cavity is along PA $\approx 36^\circ$, which is roughly perpendicular to the large-scale magnetic field in OMC-1. It is believed that the magnetic field plays an important role when a star is forming through gravitational collapse, with the system axis aligning with the local magnetic field^{20–22}. The case of BN is quite different. In addition, the disk orientation rules out BN as the powering source for the large-scale energetic H₂ outburst¹⁶ in the region, whose orientation is

roughly orthogonal. The misalignment between the system and the local magnetic field, and the survival of the disk in such a turbulent region suggest that BN is not associated with other embedded protostars within OMC-1, such as IRc2.

Two competitive scenarios have been proposed to account for massive star formation. One proposes a process similar to that for low-mass star formation with the massive stars formed through gravitational collapse of the molecular clouds and subsequent accretion from the circumstellar disk². An argument against this model is that the enormous radiation produced by more massive stars would halt the mass accretion, preventing very massive stars from being formed. Although ways of overcoming this have been proposed, for example, by changing the dust properties and the accretion rate^{23,24} or by outflows²⁵, it is essential to demonstrate that massive stars can be produced through accretion by observing an outflow/disk system. The mass of BN is estimated to be $\sim 7M_\odot$ or higher, and so our observations provide evidence that stars with masses of up to at least $\sim 7M_\odot$ can be formed through gravitational collapse and mass

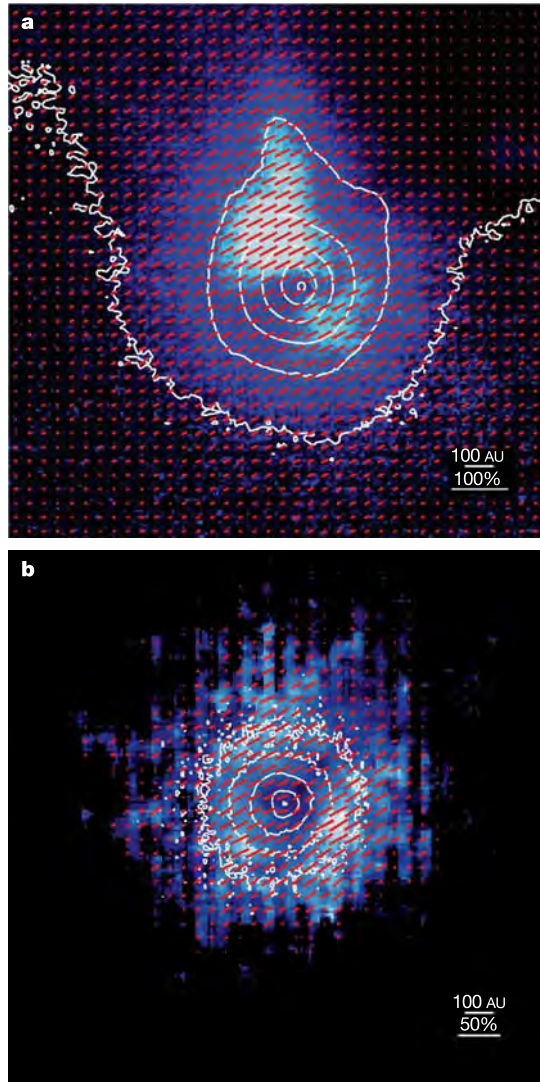


Figure 1 | Polarimetric result of the observation. **a**, Polarization vectors (red) and brightness contours (white) superposed on polarization degree image (blue) in the H-band. The contours start from $2.4 \times 10^{-4} \text{ mJy pixel}^{-1}$ with an increasing factor of $10^{1/2}$. The peak intensity is at right ascension, RA = 05 h 35 min 14.12 s, and declination, dec. = $-5^\circ 22' 23.2''$ (J2000). Polarization vectors are binned within 5×5 pixels to avoid crowding. A butterfly pattern with a dark lane (lower polarization degrees) passing through the centre of the object is clearly seen. The dark lane is approximately 500 AU ($1 \text{ AU} = 1.5 \times 10^{11} \text{ m}$) long and 100 AU wide at the narrowest section. The polarization degrees are typically 20% in the dark lane, rising to as much as $\sim 35\%$ in the two wings of the butterfly. The linear and polarization scales are shown at the lower-right corner. North is up and east is to the left. **b**, Same as **a**, but in the K-band. The contours start from $0.21 \text{ mJy pixel}^{-1}$ with an increase factor the same as that of the H-band. To improve the signal to noise ratio, the polarization image has been processed with median filtering within a box size of 3×3 pixels.

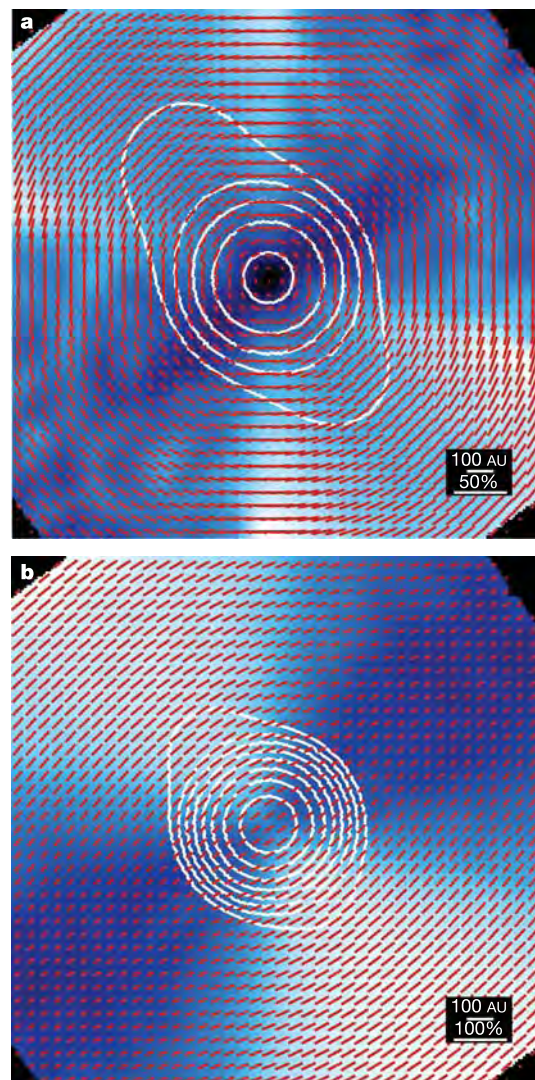


Figure 2 | Results of Monte Carlo simulations. The polarization vectors and brightness contours are superposed on the polarization degree image as in Fig. 1. The image as well as the polarization vectors is rotated to align the direction of Fig. 1. The scales of polarization degrees and linear size of 100 AU are indicated at the lower-right corner of the figure. **a**, Result of the pure scattering model; **b**, Result of the scattering model plus dichroic extinction (producing 20% polarization).

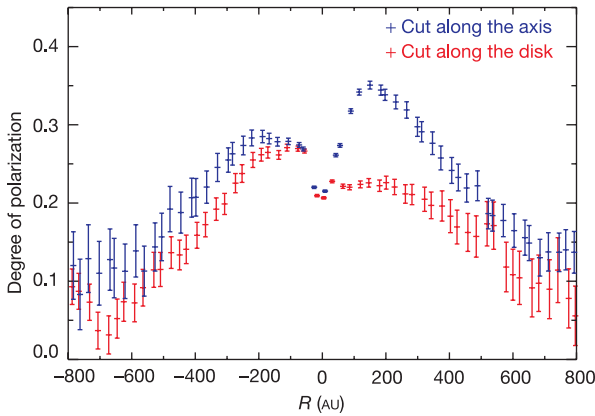


Figure 3 | Polarization degrees as a function of the distance R from the central object in the H-band. Blue dots present the polarization variance cutting along the outflow axis ($PA = 36^\circ$), while red dots are along the disk ($PA = 126^\circ$). For the blue dots, negative R is to the southwest and positive to the northeast. For red plots negative R is to the northwest and positive to the southeast. The error bars indicate the length of $\pm 1\sigma$ (standard deviation). The asymmetric polarization along the axis, within 400 AU, is significant. Although the errors are larger beyond 400 AU, the average polarization degrees in the northeastern lobe are still higher than those in the southwestern one. There is also some asymmetry in the polarization along the disk.

Table 1 | Parameters used in Monte Carlo simulation

Number of photons	10,000,000
Inner radius of accretion disk, r_0 (AU)	0.80
Outer radius of accretion disk (AU)	800
Disk radial density power law index, α	1.875
Disk flaring power law index, β	1.125
Scale height of accretion disk at $r = r_0$ (AU)	0.33
Albedo of dust grain mixture	0.50
Polar viewing angle (degrees)	85
Adopted dust model	Coated silicates ¹⁸
Inner radius of cavity (AU)	40
Conical cavity opening angle, axis to edge (degrees)	75
Cavity optical depth	0.3
System radius (AU)	1,500

accretion. Recent near-infrared spectroscopy shows some accretion signatures from massive young stars²⁶, but high-resolution near-infrared imaging polarimetry, as reported here, can provide the most direct evidence for circumstellar disks around the most massive stars.

METHODS

Imaging polarimetry. The polarimetric data were obtained on 3 January 2003 with the CIAO (Coronagraphic Imager with Adaptive Optics; ref. 27) and its polarimeter²⁸ mounted on the Subaru telescope. The pixel scale was $0.021''$. No occulting masks were used in these observations. The polarimeter consists of a stepped half-wave plate (images taken at waveplate angles of 0, 22.5, 45 and 67.5 degrees), upstream of the adaptive optics system, and a cooled wiregrid polarizer inside the CIAO cryostat. The integration time per waveplate position is 240 s at H-band and 3 s at K-band. A nearby optical star was used for wavefront sensing of the adaptive optics. The sky was clear, and the natural seeing size was $0.5''$. The spatial resolution achieved with the adaptive optics system was $0.10''$ (FWHM). After image calibrations in the standard manner, including dark subtraction, flat-fielding with dome-flats, bad-pixel substitution, and sky subtraction, using the data reduction package Image Reduction and Analysis Facility (IRAF), the Stokes parameter images can be obtained using $Q = I_0 - I_{45}$, $U = I_{22.5} - I_{67.5}$, and $I = (I_0 + I_{22.5} + I_{45} + I_{67.5})/2$. The polarization degree image is then derived by $I_p = \sqrt{Q^2 + U^2}/I$ and polarization angle image by $I_\theta = 1/2 \arctan(U/Q)$. The polarization angles are calibrated by BN itself, noted in ref. 14. Aperture polarimetry towards BN with an aperture size of $\sim 0.5''$ shows that the overall polarization degrees, 14.5% at K and 26.3% at H, agree well with the values in ref. 14.

Modelling. The modelling is based on the work of ref. 18. The detailed

description of the code and of the parameters can be found there. We used the forced scattering code to ensure calculation efficiency. The density profile of the disk is in the form:

$$\rho(r, h) = \rho_0(r/r_0)^{-\alpha} \exp\{-1/2(z/h)^2\}$$

where h is the disk scale height at radius r : $h(r) = h_0(r/r_0)^\beta$, and r_0 is the inner radius of the disk. The density profile of the envelope is in the form:

$$\rho(R, h) = C/R^{1.5} [1/((|z|/R)^{0.5} + 0.05)] \text{ for } R_{\text{cav}} < r < R_{\text{sys}}$$

$$\text{and } \rho(R, h) = \text{constant for } r < R_{\text{cav}}$$

where $R = \sqrt{r^2 + z^2}$, C is a free parameter governing the optical depth, and R_{cav} and R_{sys} are the radii of the cavity and system envelope, respectively. A dust model of silicate coated with ice has been used in the simulation. The maximum size of the particles in the dust mixture is $1.0 \mu\text{m}$. The simulation is in the H-band. Assuming an edge-on viewing angle, virtually all parameter sets give a polarization image similar to that observed, with the best-fit parameters given in Table 1. The parameter set results in an optical depth of about 8.5 at H. With the dust configuration we roughly estimate the mass of the envelope and in the cavity to be $1.7 \times 10^{-3} M_\odot$ and $1.0 \times 10^{-3} M_\odot$, respectively. The disk mass cannot be estimated from the data.

We simulate the dichroic extinction by introducing a virtual partial polarizer, producing $p\%$ polarization. The emerging Stokes parameters can be obtained by:

$$I' = 0.5[(P_1^2 + P_2^2)I + (P_2^2 - P_1^2)Q]$$

$$Q' = 0.5[(P_2^2 - P_1^2)I + (P_1^2 + P_2^2)Q]$$

$$U' = -P_1 P_2 U$$

where P_1 , P_2 are the transparencies in the x and y directions. To ensure that the polarization degree of the emerging light is $p\%$, P_1 and P_2 are set to $P_1 = [(100 + p)/(100 - p)]^{1/4} \exp(-\tau/2)$, and $P_2 = [(100 - p)/(100 + p)]^{1/4} \exp(-\tau/2)$ where τ is the optical depth along the line of sight. In the simulation the τ value does affect the polarization.

Received 21 May; accepted 29 June 2005.

- Bonnell, I. A., Bate, M. R. & Zinnecker, H. On the formation of massive stars. *Mon. Not. R. Astron. Soc.* **298**, 93–102 (1998).
- Shu, F. H., Adams, F. C. & Lizano, S. Star formation in molecular clouds—Observation and theory. *Annu. Rev. Astron. Astrophys.* **25**, 23–81 (1987).
- O'Dell, C. R. & Wen, Z. Postrefurbishment mission Hubble Space Telescope images of the core of the Orion Nebula: Proplyds, Herbig-Haro objects, and measurements of a circumstellar disk. *Astrophys. J.* **435**, 194–202 (1994).
- Perrin, M. D. *et al.* Laser guide star adaptive optics imaging polarimetry of Herbig Ae/Be stars. *Science* **303**, 1345–1348 (2004).
- Chini, R. *et al.* The formation of a massive protostar through the disk accretion of gas. *Nature* **429**, 155–157 (2004).
- Sako, S. *et al.* No high-mass protostars in the young stellar object M17–S01. *Nature* **434**, 995–998 (2005).
- Becklin, E. E. & Neugebauer, G. Observations of an infrared star in the Orion nebula. *Astrophys. J.* **147**, 799–802 (1967).
- Genzel, R. & Stutzki, J. The Orion molecular cloud and star-forming region. *Annu. Rev. Astron. Astrophys.* **27**, 41–85 (1989).
- Gezari, D. Y., Backman, D. E. & Werner, M. W. Mid-infrared imaging of Orion BN/KL. II. Luminosity sources, extinction distribution, and the nature of IRc2. *Astrophys. J.* **509**, 283–298 (1998).
- Tan, J. C. The Becklin-Neugebauer object as a runaway B star, ejected 4000 years ago from the θ^1 Orionis C system. *Astrophys. J.* **607**, L47–L50 (2004).
- Hillenbrand, L. A., Carpenter, J. M. & Skrutskie, M. F. Periodic photometric variability in the Becklin-Neugebauer object. *Astrophys. J.* **547**, L53–L56 (2001).
- Davis, L. & Greenstein, H. L. The polarization of starlight by aligned dust grains. *Astrophys. J.* **114**, 206–240 (1951).
- Jones, T. J. Infrared polarimetry and the interstellar magnetic field. *Astrophys. J.* **346**, 728–734 (1989).
- Minchin, N. R. *et al.* Near-infrared imaging polarimetry of bipolar nebulae—I. The BN-KL region of OMC-1. *Mon. Not. R. Astron. Soc.* **248**, 715–729 (1991).
- Chrysostomou, A., Hough, J. H., Burton, M. G. & Tamura, M. Twisting magnetic fields in the core region of OMC-1. *Mon. Not. R. Astron. Soc.* **268**, 325–334 (1994).
- Schultz, A. S. *et al.* NICMOS narrowband images of OMC-1. *Astrophys. J.* **511**, 282–288 (1999).
- Shuping, R. Y. & Morris, M. A new mid-infrared map of the BN/KL region using the Keck telescope. *Astron. J.* **128**, 363–374 (2004).
- Lucas, P. W. *et al.* High-resolution imaging polarimetry of HL Tau and magnetic field structure. *Mon. Not. R. Astron. Soc.* **352**, 1347–1364 (2004).
- Hough, J. H. *et al.* Infrared polarization in OMC-1—discovery of a molecular hydrogen reflection nebula. *Mon. Not. R. Astron. Soc.* **222**, 629–644 (1986).

20. Cohen, R. J., Rowland, P. R. & Blair, M. M. The source of bipolar outflow in Cepheus A. *Mon. Not. R. Astron. Soc.* **210**, 425–438 (1984).
21. Tamura, M. *et al.* Infrared polarimetry of dark clouds—III. The relationship between the magnetic field and star formation in the NGC 1333 region. *Mon. Not. R. Astron. Soc.* **231**, 445–453 (1988).
22. Matsumoto, T. & Tomisaka, K. Directions of outflows, disks, magnetic fields, and rotation of young stellar objects in collapsing molecular cloud cores. *Astrophys. J.* **616**, 266–282 (2004).
23. Wolfire, M. & Cassinelli, J. Conditions for the formation of massive stars. *Astrophys. J.* **319**, 850–867 (1987).
24. Yorke, H. W. & Sonnhalter, C. On the formation of massive stars. *Astrophys. J.* **569**, 846–862 (2002).
25. Krumholz, M. R., McKee, C. F. & Klein, R. I. How protostellar outflows help massive stars form. *Astrophys. J.* **618**, L33–L36 (2005).
26. Blum, R. D. *et al.* Accretion signatures from massive young stellar objects. *Astrophys. J.* **617**, 1167–1176 (2004).
27. Tamura, M. *et al.* Coronagraph imager with adaptive optics (CIAO): description and first results. *Proc. SPIE* **4008**, 1153–1161 (2000).
28. Tamura, M. *et al.* Near-infrared polarimeter for the Subaru telescope. *Proc. SPIE* **4843**, 190–195 (2003).

Acknowledgements This paper is based on data collected at Subaru Telescope, which is operated by the National Astronomical Observatory of Japan. This work is supported by a Grant-in-Aid from MEXT, Japan, and NSFC of China.

Author Contributions M.T., M.F., H.S. and M.I. collected the data. P.L. and M.F. did the modelling. M.T., J.H., P.L. and J.Y. contributed to the scientific discussion. Z.J. conducted data reduction and wrote the paper with help from all co-authors.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to Z.J. (zbjiang@pmo.ac.cn).

LETTERS

A photonic quantum information interface

S. Tanzilli¹, W. Tittel¹, M. Halder¹, O. Alibart², P. Baldi², N. Gisin¹ & H. Zbinden¹

Quantum communication requires the transfer of quantum states¹, or quantum bits of information (qubits), from one place to another. From a fundamental perspective, this allows the distribution of entanglement and the demonstration of quantum non-locality over significant distances^{2–6}. Within the context of applications, quantum cryptography offers a provably secure way to establish a confidential key between distant partners⁷. Photons represent the natural flying qubit carriers for quantum communication, and the presence of telecommunications optical fibres makes the wavelengths of 1,310 nm and 1,550 nm particularly suitable for distribution over long distances. However, qubits encoded into alkaline atoms that absorb and emit at wavelengths around 800 nm have been considered for the storage and processing of quantum information^{8,9}. Hence, future quantum information networks made of telecommunications channels and alkaline memories will require interfaces that enable qubit transfers between these useful wavelengths, while preserving quantum coherence and entanglement^{9–11}. Here we report a demonstration of qubit transfer between photons of wavelength 1,310 nm and 710 nm. The mechanism is a nonlinear up-conversion process, with a success probability of greater than 5 per cent. In the event of a successful qubit transfer, we observe strong two-photon interference between the 710 nm photon and a third photon at 1,550 nm, initially entangled with the 1,310 nm photon, although they never directly interacted. The corresponding fidelity is higher than 98 per cent.

Superposition of quantum states and entanglement are the fundamental resources of quantum communication and quantum information processing¹. From an abstract point of view, the nature of the carrying particle is irrelevant since only amplitudes and relative phases are exploited to encode the elementary qubits. Historically, experiments have proved many times that the fascinating properties of quantum correlations can be observed with pairs of photons¹, trapped ions^{12–14}, trapped atoms¹⁵ or cold gases¹⁶. However, the most appropriate carrier and associated encoding observable depend on the specific task. Photons have been proved suitable to transmit quantum information^{1–7}, and atoms or ions to store¹⁷ and process^{13,14} it. Photonic entanglement often relies on polarization^{4–6,18–20}, energy-time², or time-bin³ coding. Depending on the quantum communication channel, the wavelength of the photonic carrier is also important. The use of telecommunications wavelengths (1,310 and 1,550 nm) is particularly advantageous when employing optical fibres^{2,3}, while free space transmission is mostly based on shorter wavelengths^{5,6}.

Future realizations of quantum networks, containing elementary quantum processors and memories, connected by communication channels, require quantum interfaces capable of transferring qubit states from one type of carrier to another. This demands the reversible mapping between photons and atoms, which also includes the mapping between photons of different wavelengths⁹. However, as opposed to the reproduction of classical information between different media, it is not possible to merely measure the properties

of a given quantum system and replicate them accordingly, as a result of the no-cloning theorem²¹. Nevertheless, it is possible to resort to a transfer of the quantum information based on an interaction that maintains the coherence properties of the initial quantum system.

In this Letter, we demonstrate a direct quantum interface for photonic qubits at different wavelengths. As illustrated in Fig. 1, an arbitrary qubit carried by a flying telecommunications photon at 1,310 nm (λ_B), initially entangled with a photon at 1,550 nm (λ_A), is coherently transferred to another photon at a wavelength of 710 nm ($\lambda_{B'}$) using sum frequency generation (SFG). The final wavelength is close to that of alkaline atomic transitions.

In the following, we first present the theoretical description of our quantum interface. We then recall how to create and characterize maximally energy-time entangled photon pairs. The coherent quantum information transfer, taking advantage of an up-conversion stage, is then introduced. Finally, we show how this operation preserves the initial entanglement by measuring the two-photon interference between the 710 and 1,550 nm photons, and thus we demonstrate a universal qubit transfer.

Consider three systems labelled A (held by Alice), B and B' (held by Bob). In our case these systems are modes of the electro-magnetic field in optical fibres. Initially, B' is in the vacuum state denoted $|0\rangle$, while A and B may be entangled, with Schmidt coefficients c_1 and c_2 :

$$|\Psi\rangle_m = (c_1|\alpha_1\rangle_A \otimes |\beta_1\rangle_B + c_2|\alpha_2\rangle_A \otimes |\beta_2\rangle_B) \otimes |0\rangle_{B'} \quad (1)$$

The pairs of orthogonal states $|\alpha_j\rangle$ and $|\beta_j\rangle$, $j \in \{1, 2\}$, span Alice and Bob's qubit space, respectively. For example, $|\alpha_j\rangle$ could represent one photon in either the vertical ($j = 1$) or horizontal ($j = 2$) polarization state, or, as in our case, one photon in the first ($j = 1$) or second ($j = 2$) time-bin state. The desired state after a successful transfer from B to B' reads:

$$|\Psi\rangle_{\text{transfer}} = |0\rangle_B \otimes (c_1|\alpha_1\rangle_A \otimes |\beta_1\rangle_{B'} + c_2|\alpha_2\rangle_A \otimes |\beta_2\rangle_{B'}) \quad (2)$$

Such a transfer is achieved by an interaction described by the effective

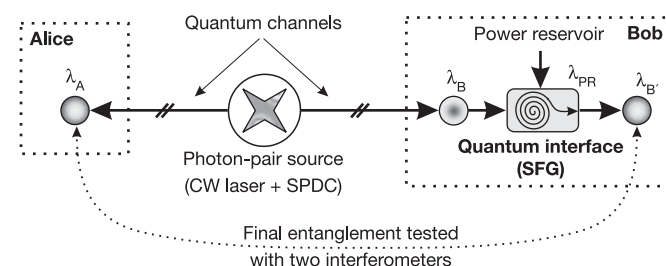


Figure 1 | Schematic illustration of the experiment concept. First, a photon-pair source produces, by spontaneous parametric down-conversion (SPDC), energy-time entangled photons at wavelengths λ_A and λ_B that are sent to Alice and Bob, respectively. Next, the qubit transfer is performed at Bob's location from photon λ_B to photon $\lambda_{B'}$ using sum frequency generation (SFG). The final entanglement between the newly created photon $\lambda_{B'}$ and Alice's photon λ_A is tested using the Franson configuration²⁵ (see Fig. 2).

¹Group of Applied Physics, University of Geneva, 1211 Geneva 4, Switzerland. ²Laboratoire de Physique de la Matière Condensée, Université de Nice-Sophia Antipolis, Parc Valrose, 06108 Nice Cedex 2, France.

hamiltonian

$$\mathcal{H} = 1_A \otimes (g_1 |0\rangle_B \langle \beta_1| \otimes |\beta_1\rangle_{B'} \langle 0| + g_2 |0\rangle_B \langle \beta_2| \otimes |\beta_2\rangle_{B'} \langle 0|) + H.c. \quad (3)$$

where g_1 and g_2 are coupling constants. The evolution can be computed (assuming an interaction time of 1 unit):

$$e^{-i\mathcal{H}}|\Psi\rangle_{\text{in}} = \cos(|g_1|)c_1|\alpha_1\rangle_A \otimes |\beta_1\rangle_B \otimes |0\rangle_{B'} - i \frac{\sin(|g_1|)}{|g_1|} g_1 c_1 |\alpha_1\rangle_A \otimes |0\rangle_B \otimes |\beta_1\rangle_{B'} \quad (4)$$

$$+ \cos(|g_2|)c_2|\alpha_2\rangle_A \otimes |\beta_2\rangle_B \otimes |0\rangle_{B'} - i \frac{\sin(|g_2|)}{|g_2|} g_2 c_2 |\alpha_2\rangle_A \otimes |0\rangle_B \otimes |\beta_2\rangle_{B'} \quad (5)$$

Note that in the second terms in equations (4) and (5), the coefficients c_j are multiplied by g_j . This implies that the transfer preserves probability amplitudes and quantum coherence if and only if the two coupling constants are equal, both in amplitude and phase: $g_1 = g_2 \equiv g$. In other words, the transfer can be achieved with perfect fidelity provided this condition is satisfied. In such a case the

evolution simplifies to:

$$e^{-i\mathcal{H}}|\Psi\rangle_{\text{in}} = \cos(|g|)|\Psi\rangle_{\text{in}} - ig \frac{\sin(|g|)}{|g|} |\Psi\rangle_{\text{transfer}} \quad (6)$$

Consequently, the transfer probability is 1 for $|g| = \pi/2$.

Figure 2 presents our two-photon source, S. It consists of a CW pump laser diode with a coherence length $\mathcal{L}_p^c$ exceeding 300 m, attenuated down to a few microwatts at $\lambda_p = 711.6$ nm, and of a quasi-phase-matched periodically poled lithium niobate waveguide^{22,23} (PPLN/W1). The poling period is chosen so as to create down-converted pairs of photons whose wavelengths are centred at 1,312 and 1,555 nm, respectively. After separation at a fibre-optic wavelength division multiplexer (WDM), the 1,555 and 1,312 nm photons are sent to Alice and Bob, respectively, using standard telecommunications optical fibres.

The long coherence length of the pump laser implies that each pump photon has a very well defined energy. Accordingly, the down-conversion process yields pairs of photons that are energy correlated as governed by energy conservation: each photon from a pair has an uncertain energy, uncertain in the usual quantum sense, but the sum of the energies of the two photons from a pair is very well defined. In

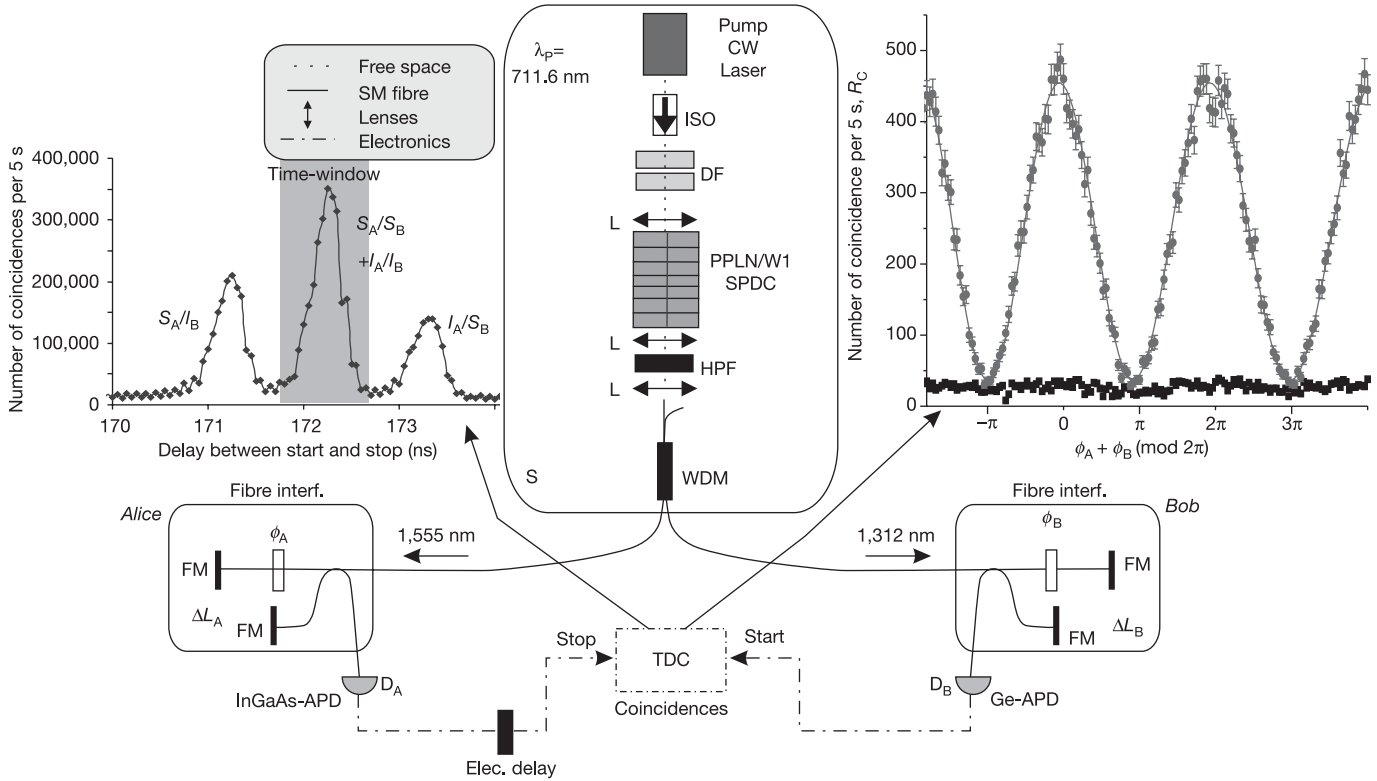


Figure 2 | Experimental Franson-type set-up used for the creation and analysis of energy-time entangled pairs of photons. Alice and Bob's analysers are equally unbalanced Michelson interferometers made of telecommunications optical fibre and Faraday mirrors (FM). The corresponding path length differences are labelled as ΔL_A and ΔL_B , respectively. The source is composed of a CW laser (Toptica Photonics DL100), an isolator (ISO), a nonlinear periodically poled lithium niobate (PPLN) waveguide, a wavelength division multiplexer (WDM) and suitable fibre coupling lenses. It also includes a high-pass wavelength filter (HPF) behind the crystal to discard the remaining pump photons, and lenses (L) to couple the created photons into a single mode (SM) fibre. The down-conversion quasi-phase matching is obtained with a specific poling period Λ of 14.1 μm and at a temperature of 85°. This 1-cm-long waveguide features a down-conversion efficiency greater than 10^{-7} . At Bob's location the germanium avalanche photodiode (Ge-APD) (NEC, D_B) is liquid nitrogen cooled and passively quenched, while the InGaAs-APD (id Quantique id200, D_A) at Alice's location is operated in gated mode. The APDs show quantum

detection efficiencies of about 10% and 14% and probabilities of dark counts of around 3×10^{-5} and 10^{-5} per nanosecond, respectively. The outputs from these APDs provide the start and stop signals for a time-to-digital converter (TDC) that records a coincidence histogram (left inset) as a function of the time difference between its two inputs. This picture is composed of three different peaks, which arise from different combinations of photon transmissions through their respective interferometer, either the short (s) or the long (l) arm. The events where both photons take the same arms (s_A/s_B and l_A/l_B) are indistinguishable as described by equation (7), leading to photon-pair interference. They can be discriminated from the other possibilities (l_A/s_B and s_A/l_B) by means of a time-resolved coincidence detection. The corresponding coincidence count rate R_C as a function of the combined phase ($\phi_A + \phi_B$) is shown in the right inset. We find sinusoidal interference fringes with more than 97% net visibility. Note finally that the variation of the combined phase is obtained by changing the temperature in Bob's interferometer.

addition, the two photons are also time correlated, as they are emitted simultaneously within their coherence time. However, the emission time of a pair is uncertain within the pump laser coherence. Hence the paired photons are entangled according to the Einstein–Podolsky–Rosen (EPR) paradox²⁴, energy and time replacing the historical position and momentum variables, respectively. More precisely, the processes of photon-pair creation at different times separated by $\Delta\tau$ are coherent as long as $\Delta\tau \ll \mathcal{L}_p^c/c$. The underlying quantum coherence of the photon pairs therefore comes from the pump laser itself.

In order to reveal and characterize the coherence carried by these energy-time entangled photon pairs we perform a Franson-type experiment²⁵ and infer the degree of entanglement from the measured two-photon interference visibility. For this purpose the two photons are sent to two analysers, one on Alice's and the other on Bob's side, that consist, in our case, of unbalanced Michelson interferometers made from standard telecommunications, fibres, fibre-optic beam-splitters, and Faraday mirrors¹ (Fig. 2). To prevent single photon interference, the optical path length difference in each interferometer (ΔL_A and ΔL_B , respectively) has to be much greater than the coherence length of the single photons, $\mathcal{L}_{s,i}^c \approx 150 \mu\text{m}$, deduced from the 15 nm spectral bandwidth of the down-converted light. Each single photon therefore has an equal probability of $\frac{1}{2}$ to exit at one of the outputs of its analyser. Moreover, to maximize two-photon interference, these analysers have to be equally unbalanced, $\Delta L_A = \Delta L_B = \Delta L$, within the coherence length of the single photons.

At the same time, however, since an entangled pair represents the quantum object to be analysed, both ΔL values ($\sim 20 \text{ cm}$ in our case) have to be much smaller than the coherence length of the pair which is given by the pump laser ($\mathcal{L}_p^c > 300 \text{ m}$).

At the output of each interferometer, the single photons are detected. Using a time-resolved coincidence detection between Alice and Bob as depicted in Fig. 2, we can post-project the initial energy-time entangled state onto a time-bin entangled state of the form:

$$|\Psi\rangle_{\text{post}} = \frac{1}{\sqrt{2}}(|s_A, s_B\rangle + e^{i(\phi_A + \phi_B)}|l_A, l_B\rangle) \quad (7)$$

where ϕ_A and ϕ_B are phases associated with the path length difference ΔL_A and ΔL_B of the related interferometers. The state given by equation (7) corresponds to the events where the down-converted photons both travel through the same arms of the interferometers, that is, s_A, s_B or l_A, l_B , these two possibilities being indistinguishable. Thus, varying the combined phase ($\phi_A + \phi_B$), we observe sinusoidal interference fringes in the coincidence rate with net (accidental coincidences discarded) and raw visibilities of $97.0 \pm 1.1\%$ and $87.4 \pm 1.1\%$, respectively (see Fig. 2). Because almost all accidental coincidences are due to detector dark counts, the purity of the created state, and hence the performance of the source, should be characterized by the net visibility. As this figure of merit is very close to the 100% predicted by quantum theory, we conclude that our source provides a state close to a pure maximally

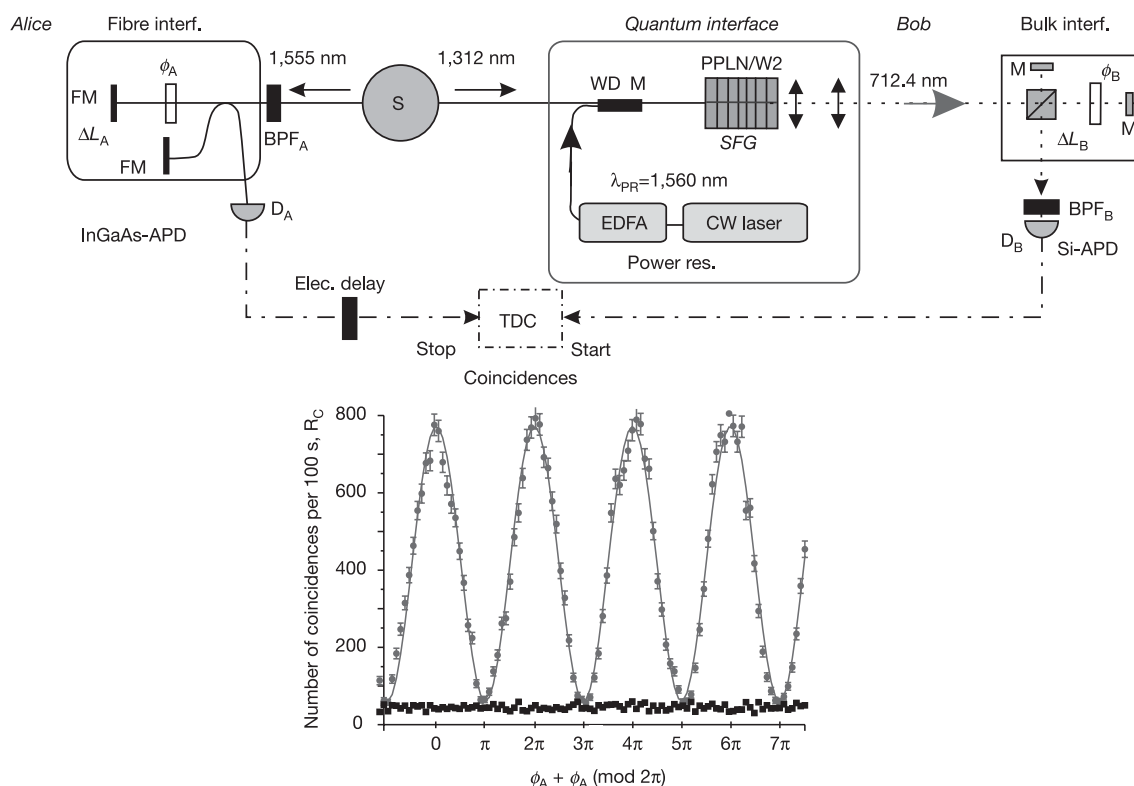


Figure 3 | Experimental set-up for the coherent transfer of quantum entanglement. On Bob's side, PPLN/W2, pumped by a high-coherence CW 700 mW power reservoir at 1,560 nm (laser 8168A from HP + EDFA from Keopsys), ensures the transfer of the qubit from 1,312 to 712.4 nm via the up-conversion process. This second 1-cm-long PPLN waveguide is of the same kind as that used for the down-conversion. They were both fabricated using the technique of soft proton exchange²², and feature almost identical phase-matching conditions. Both 1,312 and 1,560 nm wavelengths are mixed at a second WDM whose output port is directly butt-coupled to the input face of PPLN/W2 without any additional optics. At the output of the waveguide, the newly created photons at 712.4 nm are coupled into a single

mode fibre and detected using a silicon avalanche photodiode (EG&G AQ-141-FC, D_B). The coherence of the transfer is verified by measuring photon-pair interference between the 712.4 and 1,555 nm photons in the Franson configuration. The analysis on Bob's side utilizes an unbalanced, bulk-optic interferometer that is aligned with respect to Alice's. Although the coincidence rate is substantially reduced owing to a limited up-conversion probability and losses in the interferometers, the analysis of the coincidence events yields interference fringes with net and raw visibilities of 96 and 86%, respectively. This confirms the previously obtained result without the quantum interface and demonstrates, in the most general way, coherent transfer of quantum information between two photons.

entangled state, in particular, suitable to violate the Bell inequalities²⁴. Note that the missing 3% in the net visibility is most probably due to the measurement technique rather than the state preparation¹.

Now that we have characterized our entanglement resource, we proceed to the quantum information transfer, that is, the transfer of the qubit carried by Bob's particle to another photon of shorter wavelength. To this end, we replace the fibre-optic interferometer and the germanium avalanche photodiode (Ge-APD) on Bob's side by another nonlinear PPLN waveguide (PPLN/W2), a bulk-optic Michelson interferometer, and a silicon avalanche photodiode (Si-APD, D_B), as depicted in Fig 3. This second nonlinear crystal, together with a power reservoir (PR) consisting of a CW coherent laser at $\lambda_{\text{PR}} = 1,560$ nm and an erbium-doped fibre amplifier (EDFA), serves as an up-conversion stage for the incoming 1,312 nm photons produced by the previously described source, S. As a result of this operation, which is the opposite process to down-conversion, photons at 1,312 and 1,560 nm are annihilated and photons at 712.4 nm are created on Bob's side. The coherence of the PR directly relates to the phase difference between the coefficients in the effective hamiltonian of equation (3): $g_1 = \xi(t_1)$ and $g_2 = \xi(t_s)$ where $\xi(t)$ is proportional to the PR electric field at times t_1 and t_s , respectively. The quantity $c(t_1 - t_s)$ corresponds to the imbalance of the interferometers, using the notation of equation (7). The previous general condition $g_1 = g_2$ of equation (6) therefore translates, in the case at up-conversion, as a constraint on the PR's coherence: $\mathcal{L}_{\text{c}}^{\text{PR}}/c \gg t_1 - t_s$. In our experiment this condition is definitely satisfied as $\mathcal{L}_{\text{c}}^{\text{PR}} > 1$ km, which is clearly much greater than the 20 cm path length difference of Bob's analyser. Accordingly, the resulting photons at λ_{B} should become entangled with Alice photons at λ_{A} , although they never directly interacted.

The PR delivers 700 mW in a standard telecommunications fibre. We measured, with classical fields, an internal up-conversion efficiency of 80% per watt of reservoir power at optimum phase-matching^{26–28}. Note that this value is underestimated, as the losses in the waveguide itself are neglected. This up-conversion efficiency decreases by a factor of 2 when changing the single photon wavelength by ± 1 nm. Taking advantage of the energy correlation between the photons of a pair, we reduced the bandwidth by filtering the spectrum of the photons at 1,555 nm travelling to Alice to 1.5 nm (BPF_A)¹ and increased the power of the source pump laser in order to have a reasonable photon-pair production rate in this bandwidth. Hence, the overall up-conversion probability is only limited by the available pump power and coupling losses between the waveguide and single mode optical fibres at input and output faces. From this classical conversion efficiency, taking into account the reservoir power, a realistic 40% coupling efficiency into the waveguide (both for the PR and the qubit to be transferred) and the ratio between the initial and final wavelengths, we estimate the probability of a successful quantum information transfer to be: $P_{\text{success}} = 80\%W^{-1} \times 0.7W \times (0.4)^2 \times \frac{712 \text{ nm}}{1,312 \text{ nm}} \approx 5\%$. Let us mention that indirect qubit transfer can also be achieved via teleportation²⁹; however, in contrast to teleportation, quantum information transfer is achieved with a much higher success probability.

After the up-conversion stage, the 712.4 nm photons enter the temperature stabilized bulk-optic Michelson interferometer. At the output port of this analyser they are coupled, for optimal mode overlap, into a single-mode fibre adapted to visible light with a corresponding coupling efficiency greater than 60%. Then, these photons are sent to the Si-APD (D_B). In order to avoid pollution of the detector by the huge flow of PR photons at 1,560 nm, about 5×10^9 photons per ns, we use a bandpass filter centred at 712 nm (BPF_B, $\Delta\lambda = 10$ nm, more than 30 dB attenuation around 1,550 nm), as depicted in Fig. 3. We also take advantage of the combination of poor coupling and guidance of the 1,560 nm reservoir photons in the visible single mode fibre, and of the fact that Si-APDs are essentially insensitive at this wavelength. Then, the electronic signal from this APD triggers the detector of the 1,555 nm photons that arrive on

Alice's side (InGaAs-APD previously introduced and operated in gated mode). Finally, we record the coincidence events between these two detectors and measure the transfer fidelity.

Figure 3 shows the interference pattern of the coincidence events as a function of the combined phase ($\phi_{\text{A}} + \phi_{\text{B}}$) obtained with our time-resolved coincidence detection. We observe a sinusoidal oscillation with net and raw visibilities of $96.2 \pm 0.4\%$ and $86.4 \pm 0.4\%$, respectively, representing a clear signature of the preserved coherence during the quantum information transfer. The net visibility V_{net} is again very close to the 100% predicted by quantum theory. Assuming that the entire reduction of this visibility can be attributed to an imperfect quantum state transfer, we find a fidelity $F = \frac{1+V_{\text{net}}}{2}$ of 98.5%. Note however that this estimation is very conservative since the visibility observed before transfer (see Fig. 2) is hardly higher.

Note that the quantum interface demonstrated in this Letter is not limited to the specific wavelengths chosen. Indeed, suitable modifications of phase-matching conditions and pump wavelengths enable tuning to any desired wavelengths. In particular, it allows mapping of telecommunication photonic qubits onto qubits encoded at a wavelength corresponding to alkaline atomic transitions. It is particularly interesting to take advantage of periodically poled waveguiding structures, as employed here. First, phase matching conditions can easily be tuned over a broad range by changing the grating period. Second, these components yield very high up- and down-conversion efficiencies^{22,23,26}. This permits the use of a modest reservoir power to achieve a reasonable qubit transfer probability. In the case of applications requiring very narrow photon bandwidths, for instance when transferring quantum information onto atoms⁹, bright down-converters make very narrow spectral filtering possible while maintaining high photon-pair creation rates with reasonable pump powers. In addition, the narrow filtering ensures optimal phase-matching over the whole bandwidth of the photons to be up-converted, hence maximizing this process.

In this work we demonstrated, in the most general way, a direct quantum information interface between qubits carried by photons of widely different wavelengths. To this end, we verified that entanglement remains unaffected, even though one of the two entangled photons is submitted to a wavelength up-conversion process. This interface may find applications in quantum networks, where mapping of travelling qubits (for example, qubits encoded onto telecommunication photons, and stationary atomic qubits featuring resonance at much shorter wavelengths) is necessary.

Received 27 May; accepted 7 July 2005.

1. Tittel, W. & Weihs, G. Photonic entanglement for fundamental tests and quantum communication. *Quant. Inform. Comput.* **1**, 3–56 (2001).
2. Tittel, W., Brendel, J., Zbinden, H. & Gisin, N. Violation of Bell inequalities by photons more than 10 km apart. *Phys. Rev. Lett.* **81**, 3563–3566 (1998).
3. Marcikic, I. *et al.* Distribution of time-bin entangled qubits over 50 km of optical fiber. *Phys. Rev. Lett.* **93**, 180502 (2004).
4. Weihs, G., Jennewein, T., Simon, C., Weinfurter, H. & Zeilinger, A. Violation of Bell's inequality under strict Einstein locality conditions. *Phys. Rev. Lett.* **81**, 5039–5043 (1998).
5. Resch, K. J. *et al.* Distributing entanglement and single photons through an intra-city, free-space quantum channel. *Opt. Express* **13**, 202–209 (2005).
6. Peng, C.-Z. *et al.* Experimental free-space distribution of entangled photon pairs over a noisy ground atmosphere of 13 km. *Phys. Rev. Lett.* **94**, 150501 (2005).
7. Gisin, N., Ribordy, G., Tittel, W. & Zbinden, H. Quantum cryptography. *Rev. Mod. Phys.* **74**, 145–195 (2002).
8. Blinov, B. B., Moehring, D. L., Duan, L.-M. & Monroe, C. Observation of entanglement between a single trapped atom and a single photon. *Nature* **428**, 153–157 (2004).
9. Lloyd, S., Shahriar, M. S., Shapiro, J. H. & Hemmer, P. R. Long distance, unconditional teleportation of atomic states via complete Bell state measurements. *Phys. Rev. Lett.* **87**, 167903 (2001).
10. Huang, J. & Kumar, P. Observation of quantum frequency conversion. *Phys. Rev. Lett.* **68**, 2153–2156 (1992).
11. Mataloni, P., Giorgi, G. & De Martini, F. Frequency hopping in quantum interferometry. *Fortschr. Phys.* **51**, 435–441 (2003).
12. Rowe, M. A. *et al.* Experimental violation of a Bell's inequality with efficient detection. *Nature* **409**, 791–794 (2001).

13. Schmidt-Kaler, F. *et al.* Realization of the Cirac-Zoller controlled-NOT quantum gate. *Nature* **422**, 408–411 (2003).
14. Leibfried, D. *et al.* Experimental demonstration of a robust, high-fidelity geometric two ion-qubit phase gate. *Nature* **422**, 412–415 (2003).
15. Mandel, O. *et al.* Controlled collisions for multi-particle entanglement of optically trapped atoms. *Nature* **425**, 937–940 (2003).
16. Julsgaard, B., Kozhekin, A. & Polzik, E. S. Experimental long-lived entanglement of two macroscopic objects. *Nature* **413**, 400–403 (2001).
17. Julsgaard, B., Sherson, J., Cirac, J. I., Fiurásek, J. & Polzik, E. S. Experimental demonstration of quantum memory for light. *Nature* **432**, 482–486 (2004).
18. Aspect, A., Grangier, P. & Roger, G. Experimental realization of Einstein-Podolski-Rosen-Bohm *Gedankenexperiment*: a new violation of Bell's inequalities. *Phys. Rev. Lett.* **49**, 91–94 (1982).
19. Kwiat, P. G., Waks, E., White, A. G., Appelbaum, I. & Eberhard, P. H. Ultrabright source of polarization-entangled photons. *Phys. Rev. A* **60**, R773–R776 (1999).
20. Kuklewicz, C. E., Fiorentino, M., Messin, G., Wong, F. N. C. & Shapiro, J. H. High-flux source of polarization-entangled photons from a periodically poled KTiOPO₄ parametric down-converter. *Phys. Rev. A* **69**, 013807 (2004).
21. Wootters, W. K. & Zurek, W. H. A single quantum cannot be cloned. *Nature* **299**, 802–803 (1982).
22. Tanzilli, S. *et al.* PPLN waveguide for quantum communication. *Eur. Phys. J. D* **18**, 155–160 (2002).
23. Banaszek, K., U'Ren, A. B. & Walmsley, I. A. Generation of correlated photons in controlled spatial modes by downconversion in nonlinear waveguides. *Opt. Lett.* **26**, 1367–1369 (2001).
24. Aspect, A. Bell's inequality test: more ideal than ever. *Nature* **398**, 189–190 (1999).
25. Franson, J. D. Bell inequality for position and time. *Phys. Rev. Lett.* **62**, 2205–2208 (1989).
26. Roussev, R. V., Langrock, C., Kurz, J. R. & Fejer, M. M. Periodically poled lithium niobate waveguide sum-frequency generator for efficient single-photon detection at communication wavelengths. *Opt. Lett.* **29**, 1518–1520 (2004).
27. Albota, M. A. & Wong, F. N. C. Efficient single-photon counting at 1.55 μm by means of frequency upconversion. *Opt. Lett.* **29**, 1449–1451 (2004).
28. Vandevender, A. P. & Kwiat, P. J. High efficiency single photon detection via frequency up-conversion. *J. Mod. Opt.* **51**, 1433–1445 (2004).
29. Marcikic, I., De Riedmatten, H., Tittel, W., Zbinden, H. & Gisin, N. Long-distance teleportation of qubits at telecommunication wavelengths. *Nature* **421**, 509–513 (2003).

Acknowledgements We thank D.B. Ostrowsky for discussions. Financial support by the Swiss Nation Center for Quantum Photonics and the European IST project RamboQ is acknowledged. S.T. acknowledges financial support from the European Science Foundation programme 'Quantum Information Theory and Quantum Computation'.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for material should be addressed to S.T. (sebastien.tanzilli@physics.unige.ch).

A general strategy for nanocrystal synthesis

Xun Wang^{1,2}, Jing Zhuang^{1,2}, Qing Peng^{1,2} & Yadong Li^{1,2}

New strategies for materials fabrication are of fundamental importance in the advancement of science and technology^{1–12}. Organometallic^{13,14} and other organic solution phase^{15–17} synthetic routes have enabled the synthesis of functional inorganic quantum dots or nanocrystals. These nanomaterials form the building blocks for new bottom-up approaches to materials assembly for a range of uses; such materials also receive attention because of their intrinsic size-dependent properties and resulting applications^{18–21}. Here we report a unified approach to the synthesis of a large variety of nanocrystals with different chemistries and properties and with low dispersity; these include noble metal, magnetic/dielectric, semiconducting, rare-earth fluorescent, biomedical, organic optoelectronic semiconducting and conducting polymer nanoparticles. This strategy is based on a general phase transfer and separation mechanism occurring at the interfaces of the liquid, solid and solution phases present during the synthesis. We believe our methodology provides a simple and convenient route to a variety of building blocks for assembling materials with novel structure and function in nanotechnology^{13–29}.

We chose noble metals as an example to demonstrate the effectiveness of this method in yielding high quality nanocrystals. Uniform noble metal quantum dots, or nanocrystals, can be obtained through the reduction of noble metal ions by ethanol at a temperature of 20 to 200 °C under hydrothermal or atmospheric conditions. In a typical synthesis, 20 ml of aqueous solution containing noble metal salts (for example, 0.5 g of AgNO₃, HAuCl₄ or other soluble chlorides), 1.6 g sodium linoleate, 10 ml ethanol and 2 ml linoleic acid were added to a 40 ml autoclave tube under agitation. The reactions were controlled at different temperatures for specific metals, for example, 80 to 200 °C for Ag, 20 to 200 °C for Ru, Rh and Ir, 20 to 100 °C for Au, Pd and Pt. The system was sealed and treated at the designated temperature for 10 hours. After the reaction was cooled to room temperature, the products were collected at the bottom of the vessel. Based on the same synthetic process, other fatty acid and corresponding salt systems can play the same roles as the linoleic acid system.

Figure 1a shows transmission electron microscope (TEM) images of typical samples of Ag, Au, Rh and Ir nanocrystals and indicates the large quantity and good uniformity (see Supplementary Information Part I and II) that were achieved using this approach. The Ag and Au nanocrystals are usually in round shapes with smooth surfaces, and self-assemble into ordered two-dimensional (2D) arrays on the surface of the TEM grid (Fig. 1a). The diameters of the nanocrystals can be reasonably tuned from about 4 to 15 nm by altering temperature, the mole ratio of the protecting reagents to noble metal ions or the chain length of the fatty acid (see Supplementary Information Part III). This approach has also been shown to yield the nearly monodisperse ultrafine metal nanocrystals of Ru, Rh, Ir, Pd and Pt with diameters of approximately 3 nm or less. Thorough high resolution (HR) TEM characterizations revealed the highly crystalline nature of these nanocrystals. Typical HRTEM images of Ir nanocrystals with diameters ~1.7 nm show an interplanar spacing

of ~0.22 nm, which corresponds to the (111) planes of face-centred cubic Ir. EDS (energy dispersive spectroscopy) microanalysis and powder XRD (X-ray diffraction) (Fig. 2a) measurement have proven the successful synthesis of face-centred cubic structured Ag (JCPDS 4-783), Au (JCPDS 4-784), Pd (JCPDS 46-1043), Pt (JCPDS 4-802), Rh (JCPDS 5-685), Ir (JCPDS 46-1044) and hexagonal Ru (JCPDS 6-663).

The primary reaction in the preparation of noble metal nanocrystals through this liquid–solid–solution (LSS) process involved the reduction of noble metal ions by ethanol at the interfaces of metal linoleate (solid), ethanol–linoleic acid liquid phase (liquid) and water–ethanol solutions (solution) at different designated temperatures (Fig. 3). After the aqueous solution of noble metal ions, sodium linoleate (or another sodium stearate) and the mixture of linoleic acid (or another fatty acid) and ethanol were added into the vessel in order. Three phases formed in this system: sodium linoleate (solid), the liquid phase of ethanol and linoleic acid (liquid), and the water–ethanol solution containing noble metal ions (solution). A phase transfer process of the noble metal ions occurred spontaneously across the interface of sodium linoleate (solid) and the water–ethanol solution (solution) based on ion exchange, which led to the formation of noble metal linoleate and the entering of the sodium ions into the aqueous phases. Then at a designated temperature, the ethanol in the liquid and solution phases reduced the noble metal ions at the liquid–solid or solution–solid interfaces. Along with the reduction process, the *in-situ* generated linoleic acid absorbed on the surface of the noble metal nanocrystals with the alkyl chains on the outside, through which the produced metal nanocrystals will gain hydrophobic surfaces. A spontaneous phase-separation process then occurred because of the weight of the metal nanocrystals and the incompatibility between the hydrophobic surfaces and their hydrophilic surroundings, and the noble metal nanocrystals can be easily collected at the bottom of the container.

This LSS phase transfer and separation process can generate nanocrystals with a variety of properties such as, semiconducting, fluorescent, magnetic and dielectric. The phase transfer process can occur for nearly all the transitional or main group metal ions, which gives flexibility to the reactions at the interfaces (see Supplementary Information Part IV). After the phase transfer process of the metal ions from aqueous solution to the solid phase of (RCOO)_nM, under designated reaction conditions, the Mⁿ⁺ dehydrates into oxides (to yield for example, TiO₂, CuO, ZrO₂, SnO₂ or ZnO) and/or composite oxides (to yield for example, MFe₂O₄ (M represents Fe, Co, Mg, Zn or Mn) and MTiO₃ (M represents Ba or Sr) through co-precipitation). Alternatively, Mⁿ⁺ might react with other anion species such as S^{2–} (S^{2–} was supplied by Na₂S or (NH₄)₂S, to yield for example CdS, MnS, PbS, Ag₂S, CuS or ZnS), Se^{2–} (Se^{2–} was generated by the reduction of SeO₃^{2–} by N₂H₄, to yield for example CdSe or ZnSe) or F[–] (F[–] was provided from NaF or NH₄F, to yield for example YF₃, LaF₃ or NaYF₄) to yield various functional nanocrystals.

Nearly all the bandgap semiconductors can be effectively prepared through this simple LSS phase transfer and separation method, such

¹Department of Chemistry, Tsinghua University, ²National Center for Nanoscience and Nanotechnology, Beijing, 100084 China.

as TiO_2 , CuO , ZrO_2 , SnO_2 , CdS , Ag_2S , ZnS , PbS , MnS , ZnSe and CdSe . Representative TEM images of typical semiconductors of Ag_2S , PbS , ZnSe , CdSe and TiO_2 show the successful synthesis of various uniform semiconductor nanocrystals through this LSS approach (Fig. 1b, c). Two-dimensional assembly of PbS , Ag_2S , CuO , ZnSe and CdSe nanocrystals occurred spontaneously on the copper TEM grids after the evaporation of the solvents, indicating the regular shapes and narrow size distributions of these nanocrystals. Similar to the synthesis of noble metal nanocrystals, the size of the semiconductor nanocrystals can be tuned through several factors including temperature, mole ratio and the length of alkyl chains (see Supplementary Information Part III), however, for the synthesis of selenides, the temperature was controlled above 120°C to ensure the complete reduction of SeO_3^{2-} by N_2H_4 (Fig. 2b). As mentioned above, the phase transfer process and control of the reactions at the different interfaces enabled the monodispersity and variability of the semiconductor nanocrystals obtained.

By adopting bi-metal precursors in a certain mole ratio, composite oxide nanocrystals such as magnetic MFe_2O_4 (M represents Fe, Co, Mg, Zn or Mn) and dielectric MTiO_3 (M represents Ba or Sr) can be effectively prepared through co-precipitation reactions following this

LSS phase transfer and separation method. Uniform nanocrystals of magnetic spinel MFe_2O_4 could be prepared through the coprecipitation of Fe^{2+} ions and Fe^{3+} , Co^{2+} , Mg^{2+} , Mn^{2+} and Zn^{2+} ions. As shown in Fig. 1c, magnetic nanocrystals of Fe_3O_4 and CoFe_2O_4 with diameters ~ 10 nm formed 2D patterns on the TEM grids and showed good uniformity, which will be useful in biological labelling fields. In a similar way, the reaction between Ti^{4+} and Ba^{2+} and/or Sr^{2+} under strong alkali conditions can be used to prepare uniform nanocrystals of BaTiO_3 and SrTiO_3 . Typical TEM and XRD analyses are shown in Fig. 1c and Fig. 2c, respectively, and show the formation of uniform nanocrystals of tetragonal BaTiO_3 (JCPDS 74-1960) with diameters ~ 17 nm.

Our LSS phase transfer and separation approach can also be used in generating nearly monodisperse rare earth fluorescent nanocrystals with up-conversion or down-conversion emission properties. These nanocrystals can also be prepared by tuning the reaction at the interfaces of the different phases. For example, after the phase transfer process of the rare earth ions, the reaction between NaF and the $(\text{RCOO})_n\text{Ln}$ generates LnF_3 (NaYF_4 in the case of Y; the reaction between NH_4F and Y^{3+} yields YF_3), whereas the reaction between OH^- and $(\text{RCOO})_n\text{Ln}$ generates $\text{Ln}(\text{OH})_3$ nanocrystals (Fig. 2d).

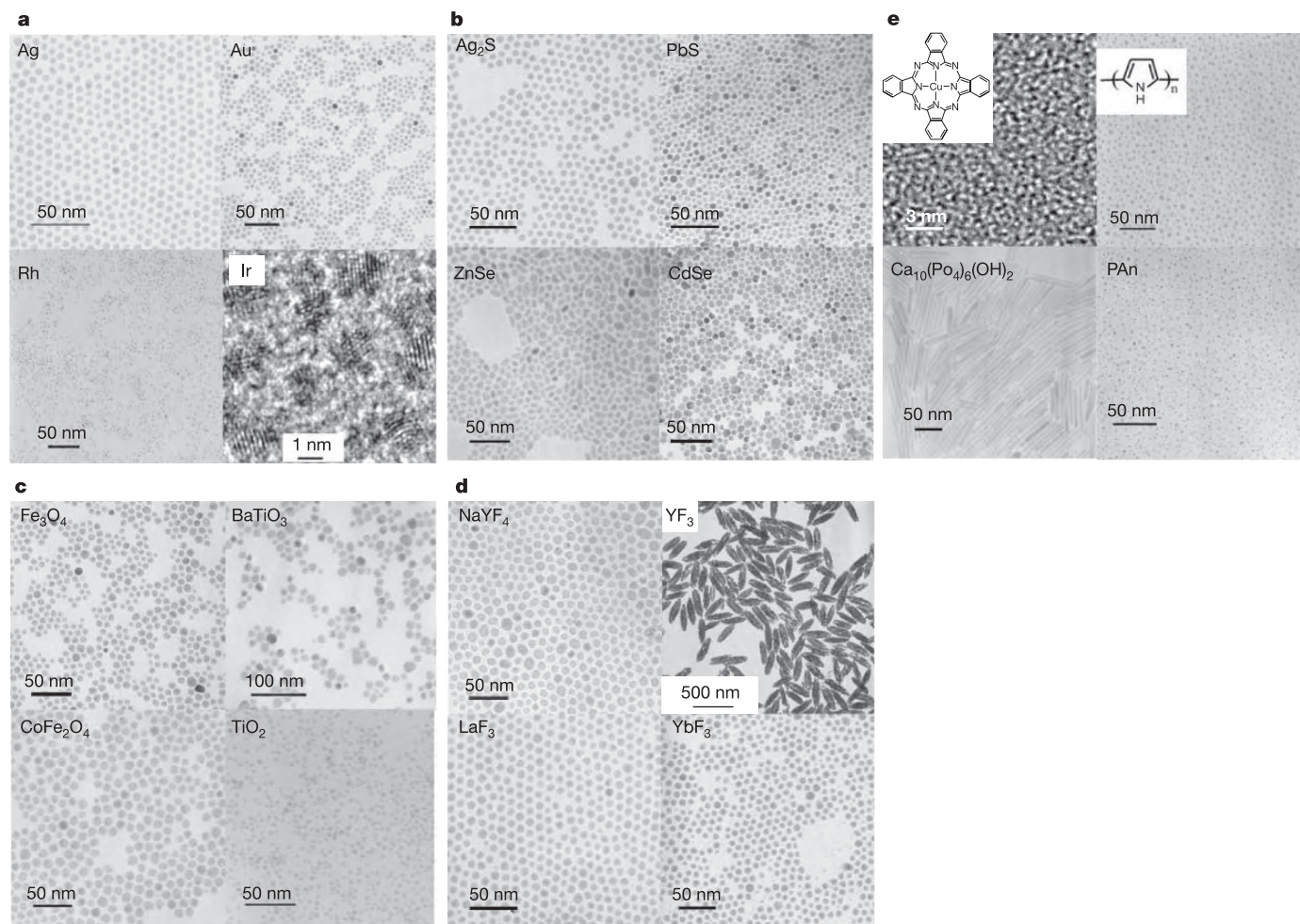
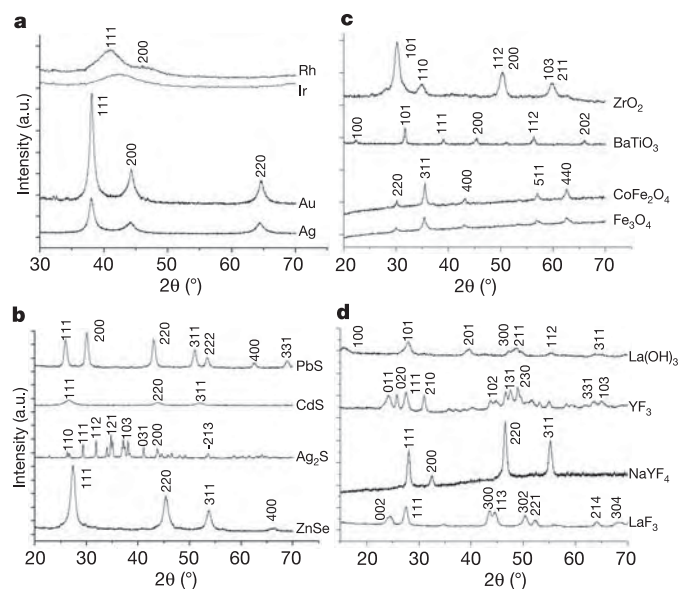


Figure 1 | TEM images of nanocrystals. **a**, Noble metal nanocrystals: Ag (6.1 ± 0.3 nm; 90°C), Au (7.1 ± 0.5 nm; 50°C), Rh (2.2 ± 0.1 nm; 120°C) and Ir (1.7 ± 0.09 nm; 120°C). **b**, Semiconductor nanocrystals: Ag_2S (7.3 ± 0.4 nm; 120°C ; $\text{Ag}^+:\text{S}^{2-}$, 2:1), PbS (5.7 ± 0.2 nm; $\text{Pb}^{2+}:\text{S}^{2-}$, 1:1), ZnSe (8.2 ± 0.9 nm; $\text{Zn}^{2+}:\text{SeO}_3^{2-}$, 1:1; 180°C) and CdSe (7.1 ± 0.8 nm; $\text{Cd}^{2+}:\text{SeO}_3^{2-}$, 1:1; 180°C). **c**, Magnetic and dielectric nanocrystals: Fe_3O_4 (9.1 ± 0.8 nm; $\text{Fe}^{2+}:\text{Fe}^{3+}$, 1:2; 160°C), CoFe_2O_4 (11.5 ± 0.6 nm; $\text{Co}^{2+}:\text{Fe}^{2+}$, 1:2; 180°C), BaTiO_3 (16.8 ± 1.7 nm; 7 g NaOH for 0.5 g $\text{Ba}(\text{NO}_3)_2$ and equal amount of TiCl_3 (in mole ratio); 180°C ; Ti^{3+} have been adopted as Ti

sources because of the relative stability of Ti^{3+} to Ti^{4+} under aqueous conditions, which will be oxidized into Ti^{4+} under hydrothermal conditions) and TiO_2 (4.3 ± 0.2 nm; 1 ml 30% TiCl_3 solution for 40 ml vessel; 180°C). **d**, Rare earth fluorescence nanocrystals: NaYF_4 (10.5 ± 0.7 nm; $\text{NaF}:\text{Y}^{3+}$, 4:1, 180°C), YF_3 ($\text{NH}_4\text{F}:\text{Y}^{3+}$, 3:1, 180°C), LaF_3 (8.0 ± 0.3 nm; $\text{F}^-:\text{La}^{3+}$, 3:1, 180°C), YbF_3 (9.5 ± 0.6 nm; $\text{F}^-:\text{Yb}^{3+}$, 3:1, 180°C). **e**, TEM images of $\text{Ca}_{10}(\text{PO}_4)_6(\text{OH})_2$, Ppy (4.2 ± 0.5 nm), PAn (3.3 ± 0.5 nm) and copper phthalocyanine (0.8 ± 0.1 nm) nanocrystals.



substrate or through a dip-coating technique, monolayer films of functional nanocrystals could be easily obtained, which will greatly increase their application in nanoscience and technology. These nanocrystals could also be re-precipitated and separated by adding an appropriate amount of ethanol to the bulky nanocrystals solutions or by force due to an external field and show advantages in processing. All these disperse/separation characteristics of the functional nanocrystals obtained through this LSS approach will provide the building blocks for the bottom-up approach to nanoscale fabrication in nanosciences and nanotechnologies.

Received 18 May; accepted 23 June 2005.

- Kroto, H. W., Heath, J. R., O'Brien, S. C., Curl, R. F. & Smalley, R. E. C60: Buckminsterfullerene. *Nature* **318**, 162–163 (1985).
- Iijima, S. Helical microtubules of graphitic carbon. *Nature* **354**, 56–58 (1991).
- Morales, A. M. & Lieber, C. M. A laser ablation method for the synthesis of crystalline semiconductor nanowires. *Science* **279**, 208–211 (1998).
- Huang, M. H. *et al.* Room-temperature ultraviolet nanowire nanolasers. *Science* **292**, 1897–1899 (2001).
- Tenne, R., Margulis, L., Genut, M. & Hodes, G. Polyhedral and cylindrical structures of Tungsten disulfide. *Nature* **360**, 444–446 (1992).
- Pan, Z. W., Dai, Z. R. & Wang, Z. L. Nanobelts of semiconducting oxides. *Science* **291**, 1947–1949 (2001).
- Duan, X. F., Huang, Y., Cui, Y., Wang, J. F. & Lieber, C. M. Indium phosphide nanowire as building blocks for nanoscale electronic and optoelectronic devices. *Nature* **409**, 66–69 (2001).
- Huang, Y., Duan, X. F., Wei, Q. Q. & Lieber, C. M. Directed assembly of one-dimensional nanostructures into functional networks. *Science* **291**, 630–633 (2001).
- Law, M. *et al.* Nanoribbon waveguides for subwavelength photonics integration. *Science* **305**, 1269–1273 (2004).
- Beck, J. S. *et al.* A new family of mesoporous molecular sieves prepared with liquid crystal templates. *J. Am. Chem. Soc.* **114**, 10834–10843 (1992).
- Huo, Q. S. *et al.* Generalized syntheses of periodic surfactant inorganic composite-materials. *Nature* **368**, 317–321 (1994).
- Yang, P. D., Zhao, D. Y., Margolese, D. I., Chmelka, B. F. & Stucky, G. D. Generalized syntheses of large-pore mesoporous metal oxides with semicrystalline frameworks. *Nature* **396**, 152–155 (1998).
- Murray, C. B., Norris, D. J. & Bawendi, M. G. Synthesis and characterization of nearly monodisperse CdE (E = sulfur, selenium, tellurium) semiconductor nanocrystallites. *J. Am. Chem. Soc.* **115**, 8706–8715 (1993).
- Peng, X. G. *et al.* Shape control of CdSe nanocrystals. *Nature* **404**, 59–61 (2000).
- Sun, S. H., Murray, C. B., Weller, D., Folks, L. & Moser, A. Monodisperse FePt nanoparticles and ferromagnetic FePt nanocrystal superlattices. *Science* **287**, 1989–1992 (2000).
- Murray, C. B., Kagan, C. R. & Bawendi, M. G. Self organization of CdSe nanocrystallites into 3-dimensional quantum-dot superlattices. *Science* **270**, 1335–1338 (1995).
- Sun, Y. G. & Xia, Y. N. Shape-controlled synthesis of gold and silver nanoparticles. *Science* **298**, 2176–2179 (2002).
- Jovin, T. M. Quantum dots finally come of age. *Nature Biotechnol.* **21**, 32–33 (2003).
- Huynh, W. U., Dittmer, J. J. & Alivisatos, A. P. Hybrid nanorod-polymer solar cells. *Science* **295**, 2425–2427 (2002).
- Tessler, N., Medvedev, V., Kazes, M., Kan, S. H. & Banin, U. Efficient near-infrared polymer nanocrystal light-emitting diodes. *Science* **295**, 1506–1508 (2002).
- Klimov, V. I. *et al.* Optical gain and stimulated emission in nanocrystal quantum dots. *Science* **290**, 314–317 (2000).
- Li, Y. D. *et al.* Bismuth nanotubes: A rational low-temperature synthetic route. *J. Am. Chem. Soc.* **123**, 9904–9905 (2001).
- Li, Y. D., Li, X. L., He, R. R., Zhu, J. & Deng, Z. X. Artificial lamellar mesostructures to WS2 nanotubes. *J. Am. Chem. Soc.* **124**, 1411–1416 (2002).
- Wang, X. & Li, Y. D. Selected-control hydrothermal synthesis of alpha- and beta-MnO2 single crystal nanowires. *J. Am. Chem. Soc.* **124**, 2880–2881 (2002).
- Wang, X. & Li, Y. D. Synthesis and characterization of lanthanide hydroxide single-crystal nanowires. *Angew. Chem. Int. Edn Engl.* **41**, 4790–4793 (2002).
- Peng, Q., Dong, Y. J. & Li, Y. D. ZnSe semiconductor hollow microspheres. *Angew. Chem. Int. Edn Engl.* **42**, 3027–3030 (2003).
- Wang, X. & Li, Y. D. Fullerene-like rare-earth nanoparticles. *Angew. Chem. Int. Edn Engl.* **42**, 3497–3500 (2003).
- Wang, X., Zhuang, J., Chen, J., Zhou, K. B. & Li, Y. D. Thermally stable silicate nanotubes. *Angew. Chem. Int. Edn Engl.* **43**, 2017–2020 (2004).
- Sun, X. M. & Li, Y. D. Ga2O3 and GaN semiconductor hollow spheres. *Angew. Chem. Int. Edn Engl.* **43**, 3827–3831 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by NSFC, the Foundation for the Author of National Excellent Doctoral Dissertation of China and the State Key Project of Fundamental Research for Nanomaterials and Nanostructures.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to L.Y. (yldli@tsinghua.edu.cn).

Modelled atmospheric temperatures and global sea levels over the past million years

Richard Bintanja¹, Roderik S.W. van de Wal¹ & Johannes Oerlemans¹

Marine records of sediment oxygen isotope compositions show that the Earth's climate has gone through a succession of glacial and interglacial periods during the past million years. But the interpretation of the oxygen isotope records is complicated because both isotope storage in ice sheets and deep-water temperature affect the recorded isotopic composition^{1–5}. Separating these two effects would require long records of either sea level or deep-ocean temperature, which are currently not available. Here we use a coupled model of the Northern Hemisphere ice sheets⁶ and ocean temperatures, forced to match an oxygen isotope record for the past million years compiled from 57 globally distributed sediment cores, to quantify both contributions simultaneously. We find that the ice-sheet contribution to the variability in oxygen isotope composition varied from ten per cent in the beginning of glacial periods to sixty per cent at glacial maxima, suggesting that strong ocean cooling preceded slow ice-sheet build-up. The model yields mutually consistent time series of continental mean surface temperatures between 40 and 80°N, ice volume and global sea level. We find that during extreme glacial stages, air temperatures were $17 \pm 1.8^\circ\text{C}$ lower than present, with a $120 \pm 10\text{ m}$ sea level equivalent of continental ice present.

Marine oxygen isotope records have provided detailed information about climatic variations over the past millions of years⁷. Interpreting the fluctuations in oxygen isotope ($\delta^{18}\text{O}$ ratio) is not straightforward, however, because the signal is affected mainly by two mechanisms (aside from local hydrographical influences). The first is preferential evaporation and subsequent incorporation of the lighter oxygen isotope in ice sheets during glacial conditions, which affects the ocean's $\delta^{18}\text{O}$ value (the 'ice-sheet part'). The second mechanism is mainly related to the uptake of $\delta^{18}\text{O}$ in calcite by benthic foraminifera, which depends on local deep-water temperature at the time of crystallization of their shells⁸ (the 'deep-water part'). Previous attempts to separate these two effects involved the use of independent temperature and sea-level records to estimate either the ice-sheet part^{1,9} or the deep-water part³. Albeit with considerable uncertainty, owing mainly to local water-mass variability, these studies have yielded long (several 100,000 yr, or 100 kyr) combined records of sea level and deep-sea temperature. These showed that the glacial deep ocean was $2\text{--}3^\circ\text{C}$ colder than today, which agrees with temperatures inferred from the Mg/Ca ratio in fossil ostracodes¹⁰.

We have tried a different approach, one that takes advantage of the fact that on glacial–interglacial timescales, the main contributors to the mean benthic oxygen isotope record—the Northern Hemisphere ice-sheet isotope content and the local deep-sea temperature—are both strongly related to Northern Hemisphere midlatitude to sub-polar surface air temperature. This puts constraints on the surface air temperature, which enabled us to separate the ice-sheet and deep-water parts simultaneously and consistently, without the need to know either in advance. Based only on $\delta^{18}\text{O}$ data, our method

additionally provides reconstructions of actual climate variables such as surface air temperature, global sea level and ice volume and ice isotope content.

Ideally, the reconstructions should be as generally representative as possible, which can be achieved by an appropriate choice of input $\delta^{18}\text{O}$ record. Many long records of benthic $\delta^{18}\text{O}$ exist today, all carrying global information, but superimposed on their global signal is the unknown effect of local hydrographical effects⁵. This variability would introduce unfavourable uncertainty in the reconstructions. For this reason, we selected a recently developed benthic $\delta^{18}\text{O}$ stack based on 57 globally distributed records¹¹, to drive our ice-sheet–ocean-temperature model (results for two individual records are shown in the Supplementary Information). In this stack, we assume

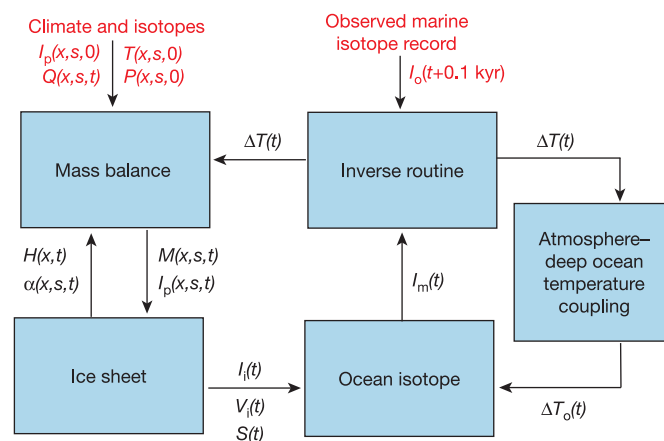


Figure 1 | Schematic outline of the inverse procedure. The routine calculates a mean surface air temperature anomaly (ΔT) based on the difference between the modelled marine isotope value ($I_m(t)$) and the observed value 0.1 kyr later ($I_o(t + 0.1 \text{ kyr})$). The value of ΔT feeds into two physical systems: (1) the ice-sheet and mass-balance module to the left and (2) the atmosphere–deep ocean temperature coupling module to the right. The mass-balance module calculates spatially (x), seasonally (s) and temporally (t) varying fields of surface mass balance (M) and isotope content of precipitation (I_p) using initial, present-day (0) fields of I_p , surface air temperature (T) and precipitation (P) and temporally varying orbitally induced insolation (Q). M and I_p are used by the ice-sheet module to calculate new distributions of ice-sheet surface height (H) and surface albedo (α), which feed back into the mass balance routine. The ice-sheet routine determines the mean ice-sheet isotope content (I_i), the ice-sheet volume (V_i) and global sea level (S). These values, together with the deep-ocean temperature (ΔT_o), are used by the ocean routine to evaluate the marine isotope value (I_m) that feeds into the inverse routine. This procedure yields mutually consistent time series of ΔT , ΔT_o , V_i , S , and I_i . Observed (input) variables are in red, modelled ones in black.

¹Institute for Marine and Atmospheric Research Utrecht, Utrecht University, Princetonplein 5, 3584 CC Utrecht, The Netherlands.

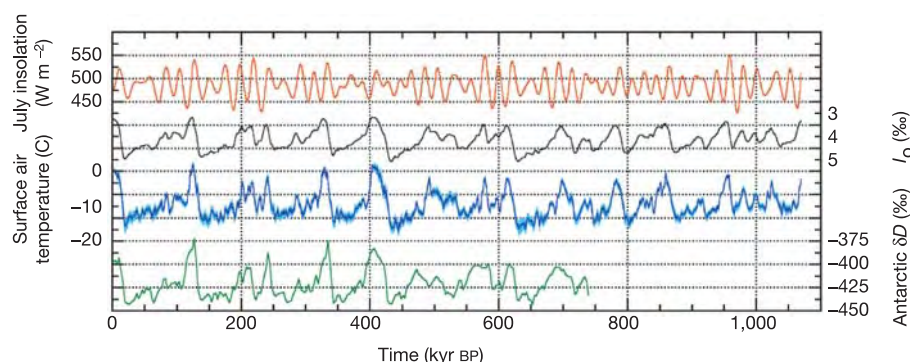


Figure 2 | 1,070-kyr time series of reconstructed Northern Hemisphere surface air temperature. Also shown are July insolation at 65° N (red), the input marine oxygen isotope stack¹¹ (black), modelled surface air temperature deviation from present (mean over the continents between 40° and 80° N) (blue), and Antarctic δD (ref. 14) (green). Insolation, $\delta^{18}O$ and

temperature are presented on the isotope stack timescale¹¹, while δD is given on its own timescale¹⁴. The error envelope for the temperature curve represents the 1 σ equivalent confidence limit derived from sensitivity tests (see Methods).

that local influences have averaged out, making it the best globally representative benthic $\delta^{18}O$ record currently available.

To separate this mean $\delta^{18}O$ signal into an ice-sheet part and a deep-water part, we used an 'inverse' technique, visualized in Fig. 1, which is based on a similar method proposed recently¹². The basic concept of this technique is to let the model determine Northern Hemisphere subpolar surface air temperature—the variable that governs isotope storage in land ice (and hence its extraction from the ocean) and deep-ocean temperatures—under the stringent requirement that the input marine $\delta^{18}O$ record is accurately followed (see Methods for details). We let the model start at 1,070 kyr before present (BP), with interglacial conditions similar to today's, and run to the present. This yielded uninterrupted and internally consistent 1,070-kyr time series of surface air temperature deviation from present (mean continental value between 40° and 80° N), ice volume, global sea level, and the ice-sheet and deep-water components of the oxygen isotope signal at 0.1-kyr time intervals.

The fluctuations in reconstructed Northern Hemisphere surface air temperature closely follow those in the input oxygen isotope record (Fig. 2). In our inverse procedure, when the ocean gets isotopically heavier, air temperatures must decrease to build isotopically light ice sheets and to cool the deep ocean, the former being a comparatively slow mechanism. The input isotope signal also dictates the amplitude of the glacial–interglacial temperature changes, with the strongest variations occurring over the past 700 kyr. The modelled extreme Northern Hemisphere surface air temperatures for this period were 17 °C below present, with a 1 σ error of 2.7 °C, in line with land-based proxy surface air temperature anomalies in subpolar Eurasia for the Last Glacial Maximum (LGM; 20 kyr BP) of –12 to –18 °C (ref. 13). The mean temperature over the 1,070-kyr period was calculated as 9.4 °C below present.

Particularly for the last 400 kyr, our reconstructed midlatitude–subpolar Northern Hemisphere temperature record compares favourably with the independent δD (deuterium) record of Dome C, Antarctica¹⁴, the longest air temperature proxy record available at present (Fig. 2). This suggests a global coherence in temperature fluctuations on multi-millennium timescales, although uncertainties in the dating of both records prevent us from reliably determining possible interhemispheric phase differences. The magnitudes of glacial–interglacial temperature changes are in line with those in δD , and warmer-than-present interglacials at 400, 330 and 120 kyr BP can be seen in both records. Our approach shows that the beginnings of glacial periods are always marked by strong temperature drops⁹ (maximum cooling of 4 °C kyr^{–1}), in agreement with the Antarctic δD record. The reconstructed cooling rate during the inception stage of the last glacial around 115 kyr BP also agrees with the $\delta^{18}O$ -inferred

temperature decrease observed in the North Greenland Ice-core Project (NGRIP) ice core, central Greenland¹⁵. Hence, the build-up of ice sheets following interglacials is associated with abrupt climate cooling. The glacial–interglacial Antarctic δD shift corresponds to a temperature amplitude of about 10 °C, suggesting that Northern Hemisphere glacial periods were significantly colder than the Antarctic, relative to the present⁴.

The fluctuations in reconstructed sea levels follow those in the input $\delta^{18}O$ signal (Fig. 3a). During the greatest glacial extremes, sea level was 125 ± 12 m below present, with less extreme lows before 700 kyr BP. Our reconstructed sea-level record agrees well with sea-level estimates based on the 470-kyr Red Sea sediment $\delta^{18}O$ record¹⁶, one of the longest continuous sea-level records available, in particular for the last glacial cycle (Fig. 3b), and also with sea-level data

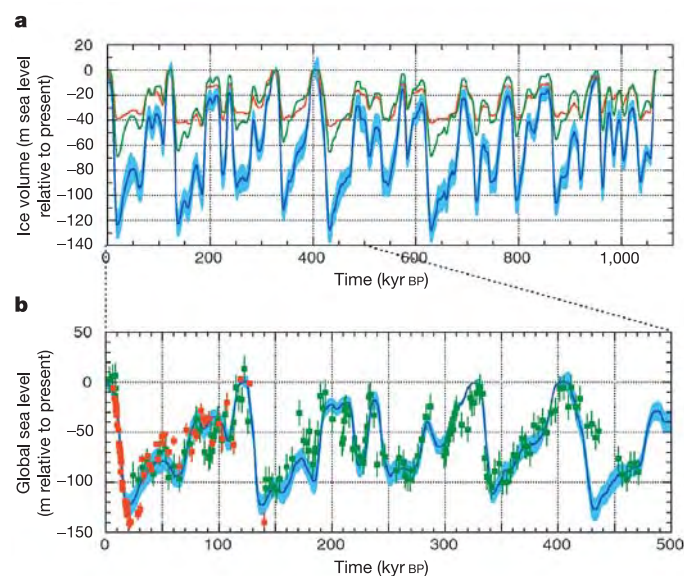


Figure 3 | Time series of past global sea level. **a**, Reconstructed global sea level (blue), with the 1 σ confidence limit based on sensitivity tests (see Methods), and the contributions from Eurasia (red) and North America (green) in sea level equivalents. Fluctuations in global sea level are dominated by the ice sheets in North America. **b**, Reconstructed global sea level and its 1 σ uncertainty (blue) compared with the 470-kyr Red Sea basin sediment $\delta^{18}O$ record¹⁶ (green squares) and New Guinea and Barbados coral reef data for the last glacial cycle¹⁷ (red squares). Timescales and error bars are taken from the respective studies.

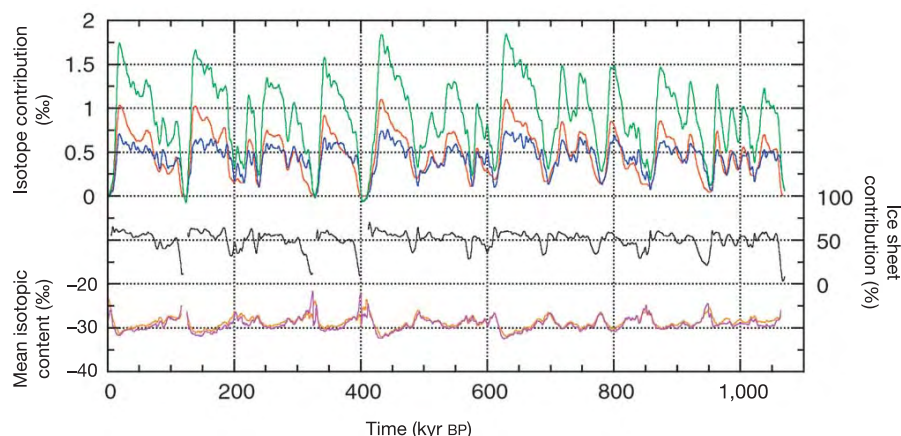


Figure 4 | Fluctuations in the various isotope signals over the 1,070-kyr period. The observed marine isotope signal with the present-day value of 3.22‰ subtracted (green) is composed of an ice-sheet contribution (red) and a deep-water temperature (blue) contribution (both simulated). The modelled relative contribution of ice sheets to the total marine isotope signal

varies strongly over glacial–interglacial cycles (black). In contrast, the simulated mean isotopic content of the ice sheets in Eurasia (orange) and North America (purple) does not exhibit large fluctuations. Missing values represent periods with no significant continental ice.

deduced from fossil coral reefs¹⁷. Interestingly, the contribution of the Eurasian ice sheets was 40 m of sea-level equivalent at most (Fig. 3a). During the most intense glacial stages, Northern Hemisphere ice-sheet growth was thus limited to North America, which is consistent with geomorphological reconstructions¹⁸ and modelling studies¹⁹, at least for the LGM. When cold glacial conditions persisted for 100-kyr periods, the Laurentide ice sheet in the east of North America and the Cordilleran ice sheet in the west were able to merge into one massive ice sheet covering much of present-day Canada¹⁹. In Eurasia, the Fennoscandian ice sheet was consistently pushed to the western continental shelf margin, whereas very dry conditions prevented expansion to the east²⁰. This geographical effect restricted further growth of the Fennoscandian ice sheet under cold glacial conditions.

The modelled mean isotopic compositions of the ice sheets only vary between -25 and -32 ‰ (Fig. 4, bottom). This indicates that the ice-sheet part of the marine isotope signal is governed by ice-volume fluctuations. For the Laurentide ice sheet, the LGM value of -32 ‰ compares very well with estimates based on $\delta^{18}\text{O}$ values in subglacial calcite of -31 ± 3 ‰ (ref. 8). The modelled deep-water part of the marine $\delta^{18}\text{O}$ signal exhibits a rapid increase at the beginning of glacials (Fig. 4, top), indicative of strong ocean cooling (2 to 2.5 °C) preceding ice-sheet build-up. This finding is consistent with sea surface temperature estimates based on other North Atlantic sediment cores²¹, and with strong cooling of the surface air as indicated by ice-core temperature proxies^{14,15} (Fig. 2). Enhanced storage of light isotopes in the expanding ice sheets causes the ice-sheet part of the $\delta^{18}\text{O}$ signal to increase steadily towards glacial extremes, followed by a sharp decline during deglaciations: it systematically rises from barely 10‰ in the beginning of glacials to about 60‰ during glacial maximums (Fig. 4, middle), indicating that marine isotope values cannot be linearly converted into sea level (see Supplementary Information). The latter translates into a mean ocean isotope increase of 1.05‰ during the LGM; this matches estimates based on $\delta^{18}\text{O}$ of relict pore water of 1.05 ± 0.2 ‰ (refs 8, 22) very well.

The main strength of our method is that it yields long and mutually consistent records of surface air temperature, ice volume and global sea level by separating the ice-sheet and deep-water parts of the marine $\delta^{18}\text{O}$ signal. This allows us to precisely compare the timing of the fluctuations of these climate variables. As an example, our results show that temperature changes (atmosphere and ocean) lead changes in Northern Hemisphere ice volume, which are dominated by the North American ice sheets, particularly during glacial

inception stages but also during terminations⁵. Future analyses may yield clues about causes and effects concerning ice age climatic fluctuations.

METHODS

Inverse procedure. The inverse procedure (observation-constrained forward modelling) centres around the calculation of surface air temperature (continental mean over $40^\circ - 80^\circ\text{N}$) relative to the present (ΔT) based on the difference between the modelled marine isotope value ($I_m(t)$) and the observed value δt ($= 0.1$ kyr) later ($I_o(t + \delta t)$): $\Delta T = \Delta T_b + c[I_m(t) - I_o(t + \delta t)]$. The first term on the right-hand side represents the average ΔT over the last b kyr that precede time t , while the parameter c in the second term determines how quickly ΔT responds to fluctuations in the marine isotope record. Optimal values of b ($= 2$ kyr) and c ($= 160^\circ\text{C}\text{‰}^{-1}$) were obtained by minimizing the sum of the squared differences between modelled and observed oceanic $\delta^{18}\text{O}$ over the 1,070-kyr period. Using these values, the difference always remained less than 0.25% of the $\delta^{18}\text{O}$ signal.

The procedure is then as follows (see also Fig. 1): if the modelled isotope value is smaller than the value of I_o 0.1 kyr later, temperatures should drop because (1) the model ocean must increase its relative heavy-isotope content through the build-up of ice sheets (where light isotopes are preferentially stored), and (2) the modelled uptake of ^{18}O in the carbonate shells of foraminifera should increase (obviously, warming occurs for $I_m > I_o$). The new marine isotope value is subsequently evaluated from the updated atmospheric air temperature according to: $V_o I_m - (V_o I_m)_{\text{PD}} = -\lambda V_o I_i - \gamma \Delta T_o$, where V_o and V_i represent ocean and ice volume (in km^3 water equivalent) respectively (PD refers to present-day conditions), I_i is the ice sheet isotope content (in ‰), ΔT_o is the deep-water temperature anomaly from present (in °C) and λ and γ are empirical parameters (discussed below). Both the ice-sheet and deep-water terms on the right-hand-side become larger when temperatures drop (I_i and ΔT_o have negative values). Hence, the simulated marine isotope value (I_m) increases as required. Its new value is then passed on to the following time step to evaluate the next change in air temperature. Starting at 1,070 kyr BP, this procedure runs to the present with 0.1-kyr steps.

Ice-sheet and deep-water models. The physical part of the method consists of two components: (1) the ice-sheet model, including an isotope budget, and (2) the atmosphere–deep-water temperature coupling. The first involves the ocean's changing isotopic composition, which depends on the amount of (light) isotopes stored in continental ice ($V_i I_i$). To simulate the evolution of the Northern Hemisphere ice sheets (excluding Greenland⁵), we used a detailed three-dimensional thermomechanical ice sheet + ice shelf + bedrock model to solve the prognostic equations for ice thickness, ice temperature and bedrock height^{6,12}. Changes in ice-sheet isotope budget were explicitly calculated, which has rarely been attempted before^{23,24}. It is assumed that spatial and temporal $\delta^{18}\text{O}$ variations in accumulating snow depend on surface air temperature^{25,26}. Once part of the ice sheet, the $\delta^{18}\text{O}$ content—a passive tracer—is transported down-slope with the vertically averaged ice velocity. The majority of the ice-sheet part of the marine oxygen isotope changes can be attributed to the Northern

Hemisphere ice sheets, but not all. On the basis of interglacial–glacial differences in isotope content and volume of the Antarctic and Greenland ice sheets^{8,27} we estimate that together they contribute 5% to global oceanic isotope variations (hence $\lambda = 1.05$), and 15% to sea-level variations⁶.

The second model component relates to the dependence of benthic $\delta^{18}\text{O}$ on deep-water temperature (ΔT_o). To take this into account, we needed to incorporate a coupling between mean deep-water temperature and Northern Hemisphere surface air temperature, which are related through vertical mixing and deep-water formation. Several proxy-based deep-water temperature reconstructions show that bottom-water temperatures were several degrees lower during glacials than during interglacials^{8–10}. Moreover, long-term (glacial–interglacial) deep-water and surface temperature proxy variations exhibit sufficient coherence to justify using a simple relation between them. Hence, we parameterized the dependence of deep-water temperature (δT_d) on the 3-kyr mean Northern Hemisphere midlatitude surface air temperature (δT_a)—which are both deviations from the present—by using a idealized climate–ocean model²⁸: $\delta T_d = 0.2\delta T_a$. This effectively introduces a lag of 2–5 kyr between δT_d and δT_a . An empirical linear relation²⁹ between deep-water temperature and $\delta^{18}\text{O}$ with a slope of $-0.28\text{‰ } ^\circ\text{C}^{-1}$ (ref. 8) then quantifies the deep-sea part of the oxygen isotope signal (hence $\gamma = 0.28\text{‰ } ^\circ\text{C}^{-1}$).

Estimation of the uncertainty. Several sensitivity tests were carried out to quantify the uncertainty associated with the method (error envelopes in Figs 2 and 3). These are based on the uncertainty in key model/input parameters¹²: the input oxygen isotope record (reflecting the standard error of the stack¹¹, a measure of local deep-water mass variability^{5,8}), changes in seasonality of the temperature forcing¹², the temperature– $\delta^{18}\text{O}$ relationship in precipitation²⁵ (related to the temporally variable and poorly quantified Dole effect³⁰, among other things), the atmosphere–deep-water temperature relationship, and the isotopic contributions from Antarctica and Greenland.

Received 7 March; accepted 29 June 2005.

- Waelbroeck, C. *et al.* Sea-level and deep water temperature changes derived from benthic foraminifera isotopic records. *Quat. Sci. Rev.* **21**, 295–305 (2002).
- Lambeck, K., Esat, T. M. & Potter, E. Links between climate and sea levels for the past three million years. *Nature* **419**, 199–206 (2002).
- Shackleton, N. J. The 100,000-year ice-age cycle identified and found to lag temperature, carbon dioxide, and orbital eccentricity. *Science* **289**, 1897–1902 (2000).
- Paillard, D. Glacial cycles: toward a new paradigm. *Rev. Geophys.* **39**, 325–346 (2001).
- Lea, D. W., Martin, P. A., Pak, D. K. & Sperbo, H. J. Reconstructing a 350-kyr history of sea level using planktonic Mg/Ca and oxygen isotope records from a Cocos Ridge core. *Quat. Sci. Rev.* **21**, 283–293 (2002).
- Bintanja, R., van de Wal, R. S. W. & Oerlemans, J. Global ice volume variations through the last glacial cycle simulated by a 3-D ice-dynamical model. *Quat. Int.* **95–96**, 11–23 (2002).
- Imbrie, J., *et al.* in *Milankovitch and Climate* (eds Berger, A. L. & Reidel, D.) 269–305 (Kluwer Academic, Boston, 1984).
- Duplessy, J.-C., Labeyrie, L. & Waelbroeck, C. Constraints on the ocean oxygen isotopic enrichment between the Last Glacial Maximum and the Holocene: paleoceanographic implications. *Quat. Sci. Rev.* **21**, 315–330 (2002).
- Cutler, K. B. *et al.* Rapid sea-level fall and deep-ocean temperature change since the last interglacial period. *Earth Planet. Sci. Lett.* **206**, 253–271 (2003).
- Dwyer, G. S. *et al.* North Atlantic deepwater temperature change during late Pliocene and late Quaternary climatic cycles. *Science* **270**, 1347–1351 (1995).
- Lisiecki, L. E. & Raymo, M. E. A Pliocene–Pleistocene stack of 57 globally distributed benthic $\delta^{18}\text{O}$ records. *Paleoceanography* **20**, doi:10.1029/2004PA001071 (2005).
- Bintanja, R., van de Wal, R. S. W. & Oerlemans, J. A new method to estimate ice age temperatures. *Clim. Dyn.* **24**, 197–211 (2005).
- Kageyama, M. *et al.* The last glacial maximum climate over Europe and western Siberia: a PMIP comparison between models and data. *Clim. Dyn.* **17**, 23–43 (2001).
- EPICA community members, Eight glacial cycles from an Antarctic ice core. *Nature* **429**, 623–628 (2004).
- North Greenland Ice Core Project members, High-resolution record of Northern Hemisphere climate extending into the last interglacial period. *Nature* **431**, 147–151 (2004).
- Siddall, M. *et al.* Sea-level fluctuations during the last glacial cycle. *Nature* **423**, 853–858 (2003).
- Lambeck, K. & Chappell, J. Sea level change through the last glacial cycle. *Science* **292**, 679–686 (2001).
- Clark, P. U., Alley, R. B. & Pollard, D. Northern hemisphere ice sheet influences on global climate change. *Science* **286**, 1104–1111 (1999).
- Tarasov, L. & Peltier, W. R. Terminating the 100 kyr ice age cycle. *J. Geophys. Res.* **102**, 21665–21693 (1997).
- Svendsen, J. I. *et al.* Late quaternary ice sheet history of northern Eurasia. *Quat. Sci. Rev.* **23**, 1229–1271 (2004).
- Cortijo, E. *et al.* Eemian cooling in the Norwegian Sea and North Atlantic ocean preceding continental ice sheet growth. *Nature* **372**, 446–449 (1994).
- Schragg, D. P., Hampt, G. & Murray, D. W. Pore fluid constraints on the temperature and oxygen isotope composition of the glacial ocean. *Science* **372**, 1930–1932 (1996).
- Mix, A. C. & Ruddiman, W. F. Oxygen-isotope analyses and Pleistocene ice volumes. *Quat. Res.* **21**, 1–20 (1984).
- Clarke, G. K. C. & Marshall, S. J. Isotopic balance of the Greenland Ice sheet: modelled concentrations of water isotopes from 30,000 BP to present. *Quat. Sci. Rev.* **21**, 419–430 (2002).
- Cuffey, K. M. *et al.* Large Arctic temperature change at the Wisconsin–Holocene glacial transition. *Science* **270**, 455–458 (1995).
- Fricke, H. C. & O’Neil, J. R. The correlation between $^{18}\text{O}/^{16}\text{O}$ ratios of meteoric water and surface temperature: its use in investigating terrestrial climate change over geologic time. *Earth Planet. Sci. Lett.* **170**, 181–196 (1999).
- Huybrechts, P. Sea-level changes at the LGM from ice-dynamic reconstructions of the Greenland and Antarctic ice sheets during glacial cycles. *Quat. Sci. Rev.* **21**, 203–231 (2002).
- Bintanja, R. & Oerlemans, J. The effect of reduced ocean overturning on the climate of the last glacial maximum. *Clim. Dyn.* **12**, 523–533 (1996).
- Shackleton, N. J. *Les Méthodes Quantitatives d’étude des Variations du Climat au Cours du Pleistocène* CNRS, Gif sur Yvette, [in English] 203–209 (1974).
- Jouzel, J., Hoffmann, G., Parrenin, F. & Waelbroeck, C. Atmospheric oxygen 18 and sea-level changes. *Quat. Sci. Rev.* **21**, 307–314 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Financial support was provided by the Netherlands Organisation of Scientific Research (NWO), in the framework of the SPINOZA award of J. Oerlemans. Constructive remarks were provided by M. Siddall and D. Dahl-Jensen.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.B. (R.Bintanja@phys.uu.nl).

Increase in tropospheric nitrogen dioxide over China observed from space

Andreas Richter¹, John P. Burrows¹, Hendrik Nüß¹, Claire Granier^{2,3,4} & Ulrike Niemeier²

Emissions from fossil fuel combustion and biomass burning reduce local air quality and affect global tropospheric chemistry. Nitrogen oxides are emitted by all combustion processes and play a key part in the photochemically induced catalytic production of ozone, which results in summer smog and has increased levels of tropospheric ozone globally¹. Release of nitrogen oxide also results in nitric acid deposition, and—at least locally—increases radiative forcing effects due to the absorption of downward propagating visible light². Nitrogen oxide concentrations in many industrialized countries are expected to decrease³, but rapid economic development has the potential to increase significantly the emissions of nitrogen oxides^{4–7} in parts of Asia. Here we present the tropospheric column amounts of nitrogen dioxide retrieved from two satellite instruments GOME^{8,9} and SCIAMACHY¹⁰ over the years 1996–2004. We find substantial reductions in nitrogen dioxide concentrations over some areas of Europe and the USA, but a highly significant increase of about 50 per cent—with an accelerating trend in annual growth rate—over the industrial areas of China, more than recent bottom-up inventories suggest⁶.

Measurements of the satellite instruments GOME and SCIAMACHY have been used to retrieve tropospheric columns of NO₂ from space^{8,11–13}, and validation of the data product used in this study has been performed¹⁴. Analysis of the satellite data has revealed the spatial and temporal distribution of tropospheric NO₂ on a global scale (for example, see refs 8, 11–13, 15 and 16). These studies have highlighted the areas of intense pollution in industrialized regions, emissions from biomass burning, soil emissions and lightning signatures.

In contrast to *in situ* data from pollution monitoring networks, which measure the concentration near the ground, the remote sensing measurements, after correction for vertical sensitivity, yield the column amount integrated over the troposphere. With the exception of aircraft and lightning, the sources of nitrogen oxides (NO_x) are located close to the surface. As a result of this NO_x source distribution and the relatively short NO₂ chemical lifetime, the tropospheric NO₂ columns observed from space are dominated by the NO₂ amount in the boundary layer.

The NO₂ columns retrieved from space can be used to improve the currently uncertain estimates of NO_x emissions. This requires knowledge of the lifetime of NO_x and of the ratio NO₂/NO_x (refs 11, 17, 18). To a first-order approximation these values are independent of NO_x concentrations¹⁷, and local changes in NO₂ columns can be assumed to be proportional to changes in local emission. The objective of this study has therefore been to investigate temporal changes in global tropospheric NO₂ patterns retrieved from GOME and SCIAMACHY measurements and thereby to infer changes in NO_x emissions from 1996 to 2004. Initially, annual averages of tropospheric NO₂ columns derived from GOME measurements

were determined on a 0.5° × 0.5° grid for the years 1996–2002, during which the instrument provided global coverage. For each grid cell, a linear regression was performed. The resulting average gradients are shown in Fig. 1. For the majority of the globe the change in NO₂ amount over the period investigated is smaller than 6 × 10¹⁴ molecules cm⁻² (~1 × 10¹⁴ molecules cm⁻² yr⁻¹), which is below the detection limit. However, significant reductions are observed over parts of Europe and the central east coast of the USA, in particular over the Ohio valley region, which has large power plants. At the same time, an increase in NO₂ is observed in the northeast of the USA as well as a large upward trend over parts of China.

To illustrate the extent and the spatial distribution of NO₂, the averages of the tropospheric NO₂ columns derived using SCIAMACHY measurements from December 2003 to November 2004 are

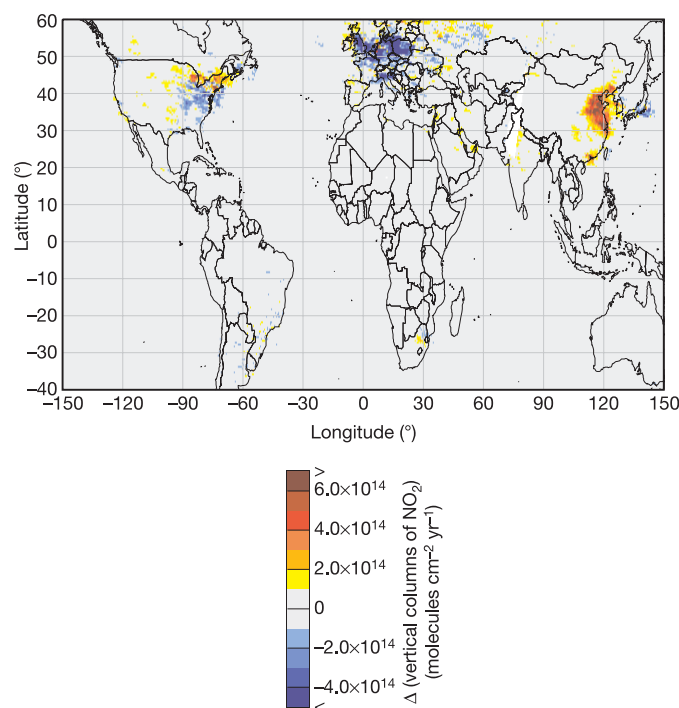


Figure 1 | Average annual changes in tropospheric NO₂ as observed by GOME from 1996 to 2002. The gradient obtained from a linear regression of the annual averages of tropospheric GOME NO₂ vertical columns, retrieved close to 10.30 a.m. LT from 1996 to 2002 is shown. Reductions in NO₂ are observed over Europe and the Central East Coast of the United States, while large increases are evident over China.

¹Institute of Environmental Physics, University of Bremen, Otto-Hahn-Allee 1, D-28359 Bremen, Germany. ²Max-Planck Institute for Meteorology, Bundesstraße 53, D-20146 Hamburg, Germany. ³Service d'Aéronomie/IPSL, University of Paris 6, Paris 75005, France. ⁴CIRES/NOAA Aeronomy Laboratory, 325 Broadway, Boulder, Colorado 80305, USA.

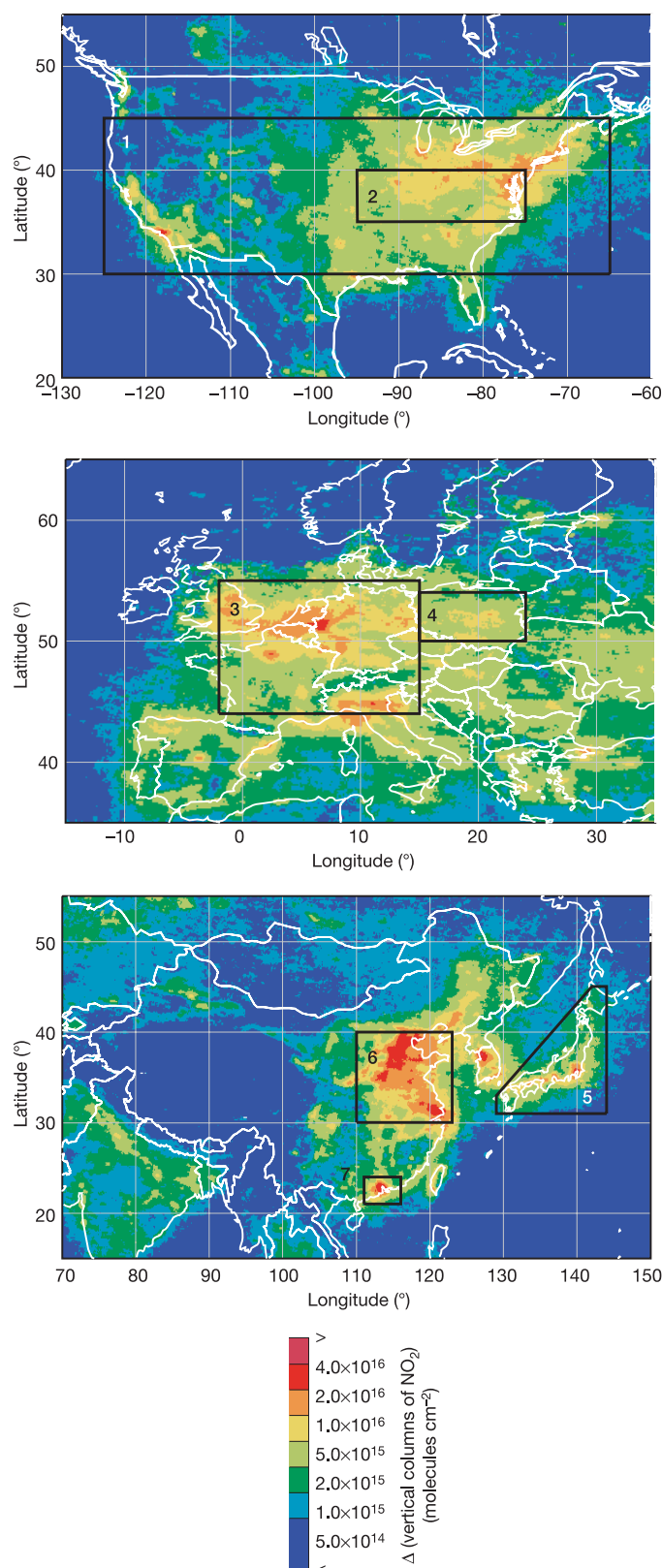


Figure 2 | SCIAMACHY tropospheric NO₂ vertical columns averaged between December 2003 and November 2004 for selected industrial regions. SCIAMACHY measurements are taken close to 10.00 a.m. LT. A nonlinear colour scale has been used because of the large range of NO₂ vertical columns. The numbered rectangles indicate the regions used in Fig. 3.

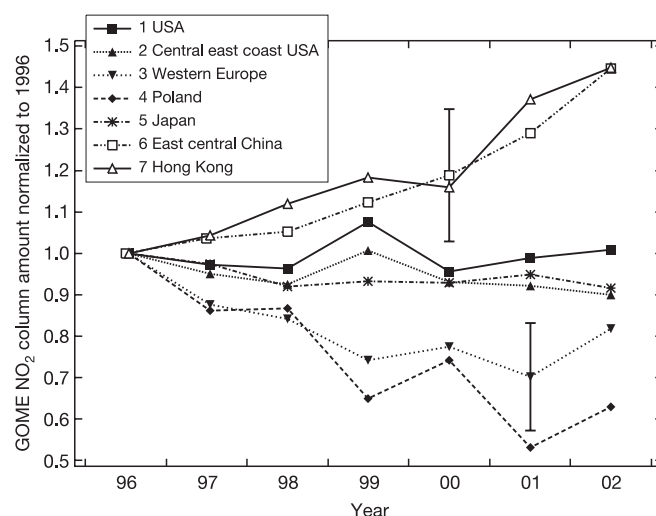


Figure 3 | The temporal evolution of tropospheric NO₂ columns from GOME for selected areas. The mean annual NO₂ column amount normalized to that in 1996 for the geographical regions USA, Central East Coast USA, Western Europe, Poland, Japan, East Central China, and Hong Kong, which are defined in Fig. 2. The error bars represent the estimated uncertainty (s.d.) for an individual year, the values over China being larger as a result of the poorer knowledge and therefore larger uncertainty of the aerosol loading and its changes.

shown in Fig. 2 for North America, Europe and Asia. As expected, the largest changes are observed in areas with large NO₂ columns. For the regions selected in Fig. 2, the temporal development of the NO₂ column is shown in Fig. 3 relative to the value measured in 1996. It is evident from these time series that the changes depicted in Fig. 1 are systematic and not dominated by year-to-year variations, in particular over Europe and China. GOME retrieval errors have been discussed in detail in several publications^{12,17,19}. Briefly, the error comprises an additive part, being approximately $(0.5-1) \times 10^{15}$ molecules cm⁻², and a multiplicative part, being 40–60% for monthly averages over polluted regions. A large part of the error budget is associated with uncertainties in the assumptions made for the radiative transfer calculations, and is systematic, because the same air-mass factors are used for each year (see Methods). Consequently only changes in these parameters contribute to the relative error, shown in Fig. 3, and the uncertainty for an annual value is reduced to about 15%.

Reductions in NO_x emissions over the last two decades have been reported for some regions. For example, a 30% NO_x emission reduction for Europe between 1990 and 2000, and 18% from 1996 to 2002 has been published³, in excellent agreement with the GOME values shown in Fig. 3. These decreases are attributed to efforts to reduce emissions by the use of catalytic converters on automobile exhaust systems, the transition to cleaner fuels, changing economic circumstances, and so on.

The evolution of the columns of NO₂ above the region of central east China (30° N, 110° E to 40° N, 123° E; see Fig. 2) are shown in more detail in Fig. 4, where monthly averages are plotted from both GOME and SCIAMACHY. A pronounced seasonal variation and a significant increase from year to year is observed, most notably in the winter values. The large NO₂ annual cycle is explained by the seasonal variation of the lifetime of NO_x in the boundary layer¹⁶, related variations in meteorological conditions, and possibly also by higher winter emissions²⁰.

The annual averages used for the linear regression in Fig. 1 are dominated by the winter values. However, if only summer months are analysed, a comparable increase of 42% (in May, June, July and August 1996–2002, $\pm 20\%$) is found, indicating a consistent increase in NO₂ concentrations over the industrialized part of China. This

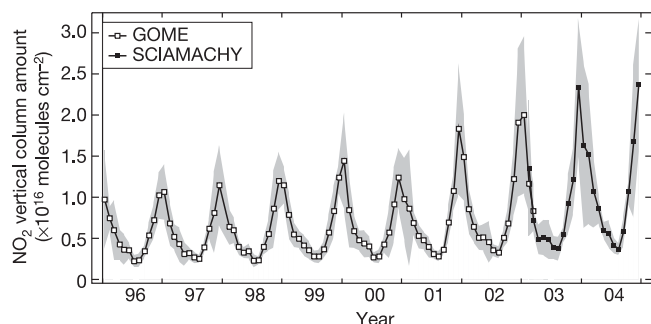


Figure 4 | Monthly averages of tropospheric vertical columns of NO₂ over East Central China. A plot of the monthly mean of the three-day composite of the tropospheric NO₂ vertical column versus time is presented for the area defined by latitudes 30° N to 40° N and longitudes 110° E to 123° E. Both GOME data (open symbols) and SCIAMACHY nadir measurements (filled symbols) are shown. SCIAMACHY nadir measurements started in August 2002, but limited data are available prior to 2003. Shaded areas represent the standard deviation estimated for the monthly mean three-day composite, and take into account the variability of the measurements resulting from changes in NO₂ and data gaps arising from cloud cover and any missing observations.

growth accelerated over the past years, from a 4% annual increase in 1997 to 12% in 2002. These GOME results are supported by recent data from SCIAMACHY, also shown in Fig. 4. The two instruments made overlapping measurements from August 2002 to June 2003, and although the measurement pattern is not identical for the two instruments (see Methods), the resulting time series fit almost seamlessly together, indicating a continuing increase in NO₂ concentrations and ruling out unidentified drift or ageing issues in the GOME data record.

To interpret the observed behaviour of retrieved NO₂ and identify its origins, the following potential effects, which might explain an increase of the observed NO₂ columns over China, should be considered.

(1) Change in measurement sensitivity. The satellite sensitivity would increase if surface reflectance increased significantly. From the GOME data there is no indication for a systematic trend in reflectance. An increase in reflecting (sulphate) aerosols would also increase sensitivity, as would a reduction in black carbon and dust aerosols¹⁷. Aerosol precursor emissions of both black carbon and sulphate in China were reported to be lower in 2000 than in 1995 (ref. 6), while TOMS (total ozone monitoring spectrometer) ultraviolet absorbing aerosol optical depths show no trend over this period²¹. Given the lack of information on aerosol trends for China, we assume that the uncertainty introduced by aerosol changes from 1996 to 2002 is of the order of the 5–10% total effect of aerosols over the USA¹⁷.

(2) Change in NO₂/NO_x partitioning. Variation of the ratio of NO₂ to NO can lead to an apparent increase of NO₂ at constant NO_x. The chemistry-transport model MOZART-2 (ref. 22) was used to quantify possible changes in the NO₂/NO_x ratio over Asia. A simulation was performed for the 1990–1999 period, where changes in surface emissions of all the chemical species included in MOZART were specified according to the POET emission scenario²³. The results of the simulations showed that, for a 60% increase in emissions, typical of the increase in the anthropogenic emissions in China, the NO₂/NO_x ratio at the overpass time of GOME increased by 8% in summer and decreased by 3% in winter. These values are significantly lower than the NO₂ changes, retrieved from GOME and SCIAMACHY.

(3) Lifetime of NO₂. This is determined primarily by the rate of production of HNO₃ from homogeneous and heterogeneous chemistry and its uptake in aerosol and clouds, and subsequent deposition. If OH concentrations are reduced, then for constant emissions, the observed NO₂ column increases. A decrease in OH

concentration can result from larger aerosol loading²⁴, but on the other hand, increasing amounts of hydrocarbons from pollution are likely to increase OH levels. In any case an OH reduction would also have large effects on the lifetime of many other atmospheric trace gases. There is no evidence from measurements for a change in OH concentration sufficient to explain our observations of the changes in the retrieved tropospheric NO₂ columns.

(4) Increase in NO₂ emissions. As result of the rapidly growing economy powered by the generation of energy from fossil fuels, large increases have been predicted for NO_x emissions from China^{4–7}. The MOZART-2 model indicates that a 60% increase in anthropogenic NO_x emissions results in a 50% increase of the NO₂ column in winter and 57% in summer over China. So to a first approximation, changes in NO₂ columns and changes in NO_x emissions are expected to be of the same order of magnitude.

Thus, a significant increase in NO_x emissions appears to be the most plausible explanation of the observed increase in NO₂ retrieved over China. In this context, a recent bottom-up inventory study⁶ concluded that NO_x emissions from China increased by 13% from 1994 to 2000, but showed signs of reducing growth rates between 1995 and 2000. Interestingly, this conclusion is not supported by GOME observations, which show a continuing increase in NO₂ columns. Apparently, any decreases in NO_x emissions for example, from improved coal-fired power stations are more than offset by other changes in emissions or sources not yet included in the inventory. The number of vehicles has doubled from 10.4 million in 1995 to 20.5 million in 2002 in China²⁵, and this in addition to increasing industrial and domestic heating sources could contribute to the observed increase of NO₂. Detailed inventory studies are needed to confirm the conclusions drawn from the satellite observations and to assign emissions to sources. The rapid increases in NO₂ observed by GOME and SCIAMACHY over the last decade demonstrate that the expanding Chinese economy has significantly increased air pollution.

METHODS

GOME and SCIAMACHY. GOME (Global Ozone Monitoring Experiment)^{8,9} is a smaller version of SCIAMACHY (Scanning Imaging Absorption spectroMeter for Atmospheric CHartography)¹⁰. Both are passive remote-sensing instruments observing the back-scattered radiance from the earth and the extraterrestrial irradiance. GOME measures in nadir whereas SCIAMACHY observes in alternate limb and nadir viewing.

The GOME is a four-channel ultraviolet/visible spectrometer observing scattered sunlight in nadir viewing geometry (ref. 8 and references therein). GOME is part of the core payload of the ESA ERS-2 platform, which was launched in April 1995 and provided global coverage from August 1995 to June 2003. The instrument observes simultaneously the spectral region from 240–793 nm having a channel-dependent spectral resolution of 0.2 to 0.4 nm. The ground scene of GOME typically has a footprint of 320 × 40 km². With an across-track swath of 960 km, global coverage at the equator is achieved within three days.

SCIAMACHY was launched in March 2002 on ENVISAT. Nadir and limb data are available since August 2002. The ultraviolet/visible nadir measurements of SCIAMACHY used here are similar to those from GOME, the two main differences being the improved spatial resolution (60 × 30 km² over most parts of the world) at reduced coverage. The latter is a result of the alternate limb nadir viewing, global coverage at the Equator being achieved in six days.

NO₂ data analysis. Satellite data are analysed for tropospheric NO₂ in a four-step procedure. First, the NO₂ absorption averaged over all light paths contributing to the signal is determined using the Differential Optical Absorption (DOAS) method in the 425–450 nm region¹². In the second step, the stratospheric component is removed by subtracting the daily stratospheric NO₂ column simulated by the 3d-CTM SLIMCAT²⁶ for the time of the satellite overpass. To account for differences between model and measurement, the SLIMCAT data are scaled to the GOME data over a clean region (180°–210° longitude). In a third step, a cloud screening is applied to remove those measurements with a cloud fraction of more than 0.2 as determined by the FRESCO algorithm²⁷.

The last step is the conversion of the tropospheric residual to a vertical tropospheric column by accounting for the vertical sensitivity of the

measurement with the radiative transfer model SCIATRAN²⁸. In this last step, a priori information is needed on surface spectral reflectance²⁹, surface altitude, aerosol loading and the shape of the vertical distribution of NO₂. The latter is taken from a run of the chemistry-transport MOZART-2 model (ref. 22) for 1997, which was used to determine monthly averaged air-mass factors on a 2.5° × 2.5° grid. Although the a priori assumptions used in the analysis have a significant impact on the retrieval results^{12,13,19}, the observed changes in NO₂ are unlikely to be affected because the same input was used for all years (see Supplementary Information). Please note that for the year 1998, no GOME data are available for January. To avoid a bias in the annual averages from the seasonal variation in NO₂, data from January 1997 were used instead.

Data analysis for GOME and SCIAMACHY is identical except that (1) SLIMCAT data are not yet available for 2004 and therefore the stratospheric correction is based on the use of a clean reference sector only, and (2) FRESKO cloud fractions are not yet available for all SCIAMACHY data and therefore an intensity criterion was used which was selected to be comparable to a threshold of 20% FRESKO cloud cover.

MOZART model. For the sensitivity studies, a MOZART-2 (ref. 22) model run at T63L47 (1.8–1.8° resolution) was used covering the time period 1990–1999. Emissions for all species for the 1990–1999 period are based on a linear interpolation for the years 1990–1995 and 1995–1997. The 1997 inventory has been compiled by combining the inventory for 1995 with regional trend data for various sources for 1995–1997. For the years 1997 to 1999 an extrapolation, based on the 1995–1997 trend, has been performed for each type of emission for each species²³. These assumptions lead to a 60% increase in anthropogenic NO_x emission over Central East China between 1990 and 1999.

Received 10 January; accepted 1 August 2005.

- Volz, A. & Kley, D. Evaluation of the Montsouris series of ozone measurements made in the nineteenth century. *Nature* **332**, 240–242 (1988).
- Solomon, S. *et al.* On the role of nitrogen dioxide in the absorption of solar radiation. *J. Geophys. Res.* **104** (D10), 12047–12058 (1999).
- Lövblad, G., Tarasón, L., Tørseth, K. & Dutchak, S. (eds) *EMEP Assessment Part I. European Perspective The Cooperative Programme for Monitoring and Evaluation of the Long-range Transmission of Air Pollutants in Europe*, October 2004 48–51 (Norwegian Meteorological Institute, Oslo, 2004).
- van Aardenne, J. A., Carmichael, G. R., Levy, H. II, Streets, D. & Hordijk, L. Anthropogenic NO_x emissions in Asia in the period 1990–2002. *Atmos. Environ.* **33**, 633–646 (1999).
- Streets, D. G. & Waldhoff, S. T. Present and future emissions of air pollutants in China: SO₂, NO_x, and CO. *Atmos. Environ.* **34**, 363–374 (2000).
- Streets, D. G. *et al.* An inventory of gaseous and primary aerosol emissions in Asia in the year 2000. *J. Geophys. Res.* **108** (D21), 8809, doi:10.1029/2002JD003093 (2003).
- Akimoto, H. Global air quality and pollution. *Science* **302**, 1716–1719 (2003).
- Burrows, J. P. *et al.* The Global Ozone Monitoring Experiment (GOME): Mission concept and first scientific results. *J. Atmos. Sci.* **56**, 151–175 (1999).
- European Space Agency, *GOME Global Ozone Measuring Experiment Users Manual* (ESA SP-1182, ESA/ESTEC, Noordwijk, 1995).
- Bovensmann, H. *et al.* SCIAMACHY—Mission objectives and measurement modes. *J. Atmos. Sci.* **56**, 127–150 (1999).
- Leue, C., Wenig, M., Wagner, T., Platt, U. & Jähne, B. Quantitative analysis of NO_x emissions from GOME satellite image sequences. *J. Geophys. Res.* **106**, 5493–5505 (2001).
- Richter, A. & Burrows, J. P. Retrieval of tropospheric NO₂ from GOME measurements. *Adv. Space Res.* **29**, 1673–1683 (2002).
- Martin, R. V. *et al.* An improved retrieval of tropospheric nitrogen dioxide from GOME. *J. Geophys. Res.* **107**, 4437–4456 (2002).
- Petriloli, A. *et al.* First comparison between ground-based and satellite-borne measurements of tropospheric nitrogen dioxide in the Po basin. *J. Geophys. Res.* **109**, D15307, doi:10.1029/2004JD004547 (2004).
- Beirle, S., Platt, U., Wenig, M. & Wagner, T. Weekly cycle of NO₂ by GOME measurements: a signature of anthropogenic sources. *Atmos. Chem. Phys.* **3**, 2225–2232 (2003).
- Kunhikrishnan, T. *et al.* Analysis of tropospheric NO_x over Asia using the model of atmospheric transport and chemistry (MATCH-MPIC) and GOME-satellite observations. *Atmos. Environ.* **38**, 581–596 (2004).
- Martin, R. V. *et al.* Global inventory of nitrogen oxide emissions constrained by space-based observations of NO₂ columns. *J. Geophys. Res.* **108**, 4537–4548 (2003).
- Richter, A. *et al.* Satellite measurements of NO₂ from international shipping emissions. *Geophys. Res. Lett.* **31**, doi:10.1029/2004GL020822 (2004).
- Boersma, K. F., Eskes, H. J. & Brinksma, E. J. Error analysis for tropospheric NO₂ retrieval from space. *J. Geophys. Res.* **109**, doi:10.1029/2003JD003962 (2004).
- Jaeglé, L., Steinberger, L., Martin, R. V. & Chance, K. Global partitioning of NO_x sources using satellite observations: Relative roles of fossil fuel combustion, biomass burning and soil emissions. *Faraday Discuss.* **130**, 407–423, doi:10.1039/b502128f (2005).
- Massie, S. T., Torres, O. & Smith, S. J. Total Ozone Mapping Spectrometer (TOMS) observations of increases in Asian aerosol in winter from 1979 to 2000. *J. Geophys. Res.* **109**, D18211, doi:10.1029/2004JD004620 (2004).
- Horowitz, L. W. *et al.* A global simulation of tropospheric ozone and related tracers: description and evaluation of MOZART, version 2. *J. Geophys. Res.* **108** (D24), 4474, doi:10.1029/2002JD002853 (2003).
- Olivier, J. *et al.* Present and Future Surface Emissions of Atmospheric Compounds. POET (Precursors of Ozone and their Effects in the Troposphere) Report no. 2, 4–12 (EU project EVK2-1999-00011, 2003).
- Martin, R. V., Jacob, D. J., Yantosca, R. M., Chin, M. & Ginoux, P. Global and regional decreases in tropospheric oxidants from photochemical effects of aerosols. *J. Geophys. Res.* **108** (D3), 4097, doi:10.1029/2002JD002622 (2003).
- Zhao, J. & Gallagher, K. S. Clean vehicle development in China. *Sinosphere J.* **6**, 20, (<http://www.chinaenvironment.net/sino/sino14.pdf>) (2003).
- Chipperfield, M. P. Multiannual simulations with a three-dimensional chemical transport model. *J. Geophys. Res.* **104**, 1781–1805 (1999).
- Koelemeijer, R. B. A., Stammes, P., Hovenier, J. W. & de Haan, J. F. A fast method for retrieval of cloud parameters using oxygen A band measurements from the Global Ozone Monitoring Experiment. *J. Geophys. Res.* **106**, 3475–3490 (2001).
- Rozanov, V., Diebel, D., Spurr, R. J. D. & Burrows, J. P. GOMETRAN: a radiative transfer model for the satellite project GOME—the plane parallel version. *J. Geophys. Res.* **102**, 16683–16695 (1997).
- Koelemeijer, R. B. A., de Haan, J. F. & Stammes, P. A database of spectral surface reflectivity in the range 335–772 nm derived from 5.5 years of GOME observations. *J. Geophys. Res.* **108** (D2), 4070, doi:10.1029/2002JD002429 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements GOME and SCIAMACHY IVO and IV1 spectra were provided by ESA through DFD/DLR. We thank M. Chipperfield for providing SLIMCAT data. This study has been funded in part by the research programmes of the University of Bremen, the Max Planck Society, the European Union, German Aerospace (DLR), the German Ministry of Science and Education (BMBF) and the European Space Agency (ESA).

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.R. (richter@iup.physik.uni-bremen.de).

Fracture surface energy of the Punchbowl fault, San Andreas system

Judith S. Chester¹, Frederick M. Chester¹ & Andreas K. Kronenberg¹

Fracture energy is a form of latent heat required to create an earthquake rupture surface and is related to parameters governing rupture propagation and processes of slip weakening^{1–3}. Fracture energy has been estimated from seismological and experimental rock deformation data^{4–8}, yet its magnitude, mechanisms of rupture surface formation and processes leading to slip weakening are not well defined^{8–10}. Here we quantify structural observations of the Punchbowl fault, a large-displacement exhumed fault^{11,12} in the San Andreas fault system, and show that the energy required to create the fracture surface area in the fault is about 300 times greater than seismological estimates would predict for a single large earthquake. If fracture energy is attributed entirely to the production of fracture surfaces, then all of the fracture surface area in the Punchbowl fault could have been produced by earthquake displacements totalling <1 km. But this would only account for a small fraction of the total energy budget, and therefore additional processes probably contributed to slip weakening during earthquake rupture.

Numerous models have been proposed that provide a basis for relating the macroscopic energy budget to physical processes of earthquake slip^{10,13}. In the commonly used slip-weakening model, the elastic strain energy released during an earthquake is partitioned between the fracture energy, E_G , the frictional heat, E_H , and the energy radiated as seismic waves, E_R (refs 10, 14) (Fig. 1). E_H , often considered a large component of the total energy budget, does not directly influence earthquake rupture dynamics, whereas the relative magnitude of E_R to E_G , expressed by radiation efficiency $\eta_R = E_R / (E_R + E_G)$, plays a fundamental role^{2,14,15}. E_G describes the flow of energy at rupture tips that is required to form a rupture surface and produce a breakdown in strength^{1,2} (Fig. 1). For small E_G ($\eta_R = 1$), an earthquake rupture is considered very brittle, rupture speed is rapid, and growth is favoured. Currently, the details of energy partitioning, the magnitude of each energy term, and whether η_R varies with earthquake magnitude and tectonic setting are not well known^{13–15}.

Some constraints on E_G are provided by laboratory estimates of the specific fracture energy, G , defined by $G = E_G / S$, where S is the area of the rupture surface¹⁰. For a tensile fracture in brittle material, G is nearly equivalent to the free-surface energy of the fracture^{16,17}. Experiments show that G for shear fracture of intact rock under pressures up to 470 MPa is greater than that for tensile cracking, and that G increases with normal stress and roughness for frictional slip^{4,5,18}. These observations may reflect the formation of microfractures and extreme fragmentation associated with the formation of shear surfaces^{4,19–22}. Some seismologic data suggest that G increases with earthquake size⁸, qualitatively consistent with scaling relations between rupture dimension and extent of associated fracture damage²². Yet whether other processes, such as fluid pressurization or melt lubrication, also contribute to rupture formation and the breakdown in strength is still unknown^{8–10,23}.

To answer this question, the magnitude of the total free-surface energy of fractures produced during an earthquake rupture must be determined. It is essential that this estimate should account for the fracture surface energy in the broad zone of damage bounding a fault, as well as that harboured within the narrow, intensely fragmented core^{2,3,7}. Accordingly, we use structural observations of the Punchbowl fault zone to quantify the total fracture surface area, S_T , which includes the surface area of cataclastic particles in the ultracataclasite, S_{UC} , the surface area of gouge particles in subsidiary faults of the damage zone, S_{SF} , and the surface area of microfractures in the damage zone, S_{MF} (Fig. 2). This study provides the first quantification of fracture surface area over the entire deformed zone of a mature fault expressed as fracture surface area per unit area (in m²) of the macroscopic fault surface. The surface area of each component is determined by direct observation of fractures and particle sizes, and scaling relations for particle size and subsidiary fault length, thickness and density. We use S_T to calculate the total fracture surface energy and relate this to the G of an earthquake. In so doing, we assume that all fracture damage in the Punchbowl fault zone was produced during seismic slip¹².

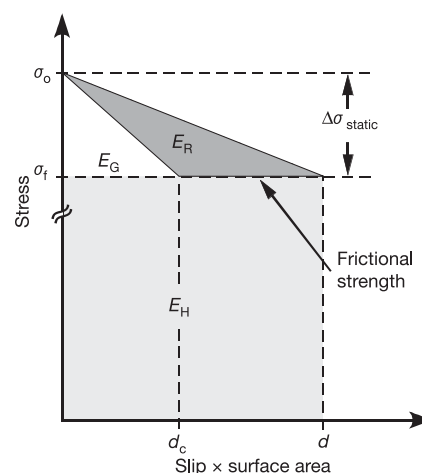


Figure 1 | Simple representation of the energy budget for earthquakes on the basis of fault slip-weakening models^{10,14,15}. The upper line bounding region E_R shows a shear stress drop from initiation of earthquake slip to a total slip, d . Frictional, or resistive, strength is shown by the lower line bounding region E_R . Frictional strength decreases with slip over a characteristic displacement, d_c , to a residual friction level. For brevity, more complex variations in shear stress and resistive strength expected for actual earthquake events^{13,15} are not shown. σ_0 , initial stress; σ_f , final, frictional stress. Details of energy partitioning are model-specific. In general, however, the radiated energy, E_R , is the difference between the mean shear stress and mean resistive strength^{2,15}.

¹Center for Tectonophysics, Department of Geology & Geophysics, Texas A&M University, College Station, Texas 77843-3115, USA.

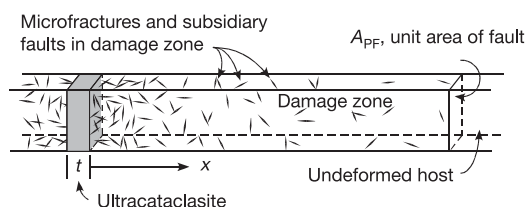


Figure 2 | Structural model of the Punchbowl fault zone for calculating fracture surface area. Domains of damaged rock containing subsidiary faults and microfractures are bounded by undeformed host rock. The central core of the fault contains an ultracataclasite layer of thickness t . Total fracture surface area, $S_T = S_{UC} + S_{SF} + S_{MF}$, is determined within the column of rock across the entire fault zone of cross-sectional area $A_{PF} = 1 \text{ m}^2$. Calculations assume damage is symmetric about the ultracataclasite layer.

The Punchbowl fault is a 200-m-thick zone of fractured rock containing a core several metres thick of sheared cataclasite and ultracataclasite^{11,12,24}. The fault records extreme localization of slip similar to other ancient exhumed and active seismic faults¹². The Punchbowl fault juxtaposes igneous and metamorphic rocks of the San Gabriel complex and arkosic sandstones of the Punchbowl Formation along a continuous layer of ultracataclasite. Deformation is dominantly brittle and the fault is exhumed to a depth of 2–4 km (ref. 11). The ultracataclasite layer is 4 cm to 1 m thick, and displays a

relatively planar, continuous surface that served as the principal slip surface during the last several kilometres of displacement¹². The principal slip surface is a layer 1 mm thick of ultracataclasite distinguished by uniform birefringence, contains porphyroclasts of older ultracataclasite that record reworking, and has at least one relatively distinct and planar boundary²⁵. Particles greater than 10 μm in diameter constitute 28% of the ultracataclasite, the majority of which are fragments of microscopic veins¹¹. Transmission electron microscopy reveals particles of host rock less than 100 nm in diameter²⁵ (Fig. 3). The smallest (4 to 50 nm) particles imaged are cataclastic particles or euhedral grains produced by syn- and post-faulting reaction. Measurements of particle size in plane section¹² are consistent with $N(D) = cD^{-a}$, where D is particle diameter, $N(D)$ is the number of particles for each size bin, and a and c are constants, with $a = 2.0$ and $c = 0.07$ (Fig. 3b).

The total fracture surface area in the ultracataclasite is determined assuming spherical particles with a power-law size distribution. Our observations indicate that the upper cut-off is 0.1 mm, and we assume that particles <50 nm in diameter follow the same power-law distribution to a lower cut-off of 1.6 nm. Using a mean layer thickness of 0.3 m and the ratio of surface area to volume for spheres ($3/r$), gives a S_{UC} of $1.3 \times 10^8 \text{ m}^2$ per unit area of the fault surface. The total surface area in the 1-mm-thick principal slip surface is $4.2 \times 10^5 \text{ m}^2$ per unit area. The surface area per unit volume of the Punchbowl ultracataclasite is similar to that determined for gouge of the San Andreas fault at Tejon Pass²⁶.

Subsidiary faults of the damage zone are consistent with scaling relations between number, length, displacement and thickness of brittle faults in crustal rock^{22,27}. Faults typically are metres in length, <1 mm thick and display centimetre offsets. Faults up to 1–2 cm thick, with 1–2-m offsets, are also present¹¹ (Fig. 4). Linear density, P_L , of subsidiary faults in the Punchbowl Formation decreases with the logarithm of distance from the ultracataclasite (Fig. 4). For the entire damage zone, we assume $N(L) = cL^{-a}$, where L is fault length, $N(L)$ is the number of faults for each length bin, $a = 1$ and $c = 2,500$, the ratio of fault displacement to length is 0.01, and the ratio of thickness to displacement is 0.005.

Subsidiary fault gouge is coarser-grained than the ultracataclasite (Fig. 3b), and particles 100–200 μm in diameter comprise 25% of the total volume. We assume particle size distributions fit $N(D) = cD^{-a}$ with constants $a = 2.0$ and $c = 0.15$, and have an upper cut-off of 200 μm and lower cut-off of 0.8 μm . Assuming the orientation of subsidiary faults is isotropic, such that the surface area is twice P_L (ref. 28), integrating over the damage zone, and using scaling relations for fault thickness and particle size, we calculate a value for S_{SF} of $3.3 \times 10^6 \text{ m}^2$ per unit area of the fault surface.

The linear density of microfractures in the Punchbowl Formation decreases with the logarithm of distance from the ultracataclasite to a constant, regional density at approximately 100 m (Fig. 4). Integrating over the damage zone, assuming microfracture orientations are isotropic, and doubling for the two surfaces of a microfracture, gives a value for S_{MF} of $2.4 \times 10^6 \text{ m}^2$ per unit fault area.

The total fracture surface area in the damage zone ($S_{SF} + S_{MF}$) is less than 10% of that in the ultracataclasite layer, giving a total fracture surface area for the entire fault zone, S_T , of $1.4 \times 10^8 \text{ m}^2$ per unit fault area. Assuming the free-surface energy to be 1 J m^{-2} , and that the geometric surface area should be increased by about a factor of five to account for non-spherical grain shapes and finer-scale roughness^{16,19,26}, brings the total fracture surface energy to $7 \times 10^8 \text{ J}$ per unit fault area. Fracture surface energy of the millimetres-thick principal slip surface is $2 \times 10^6 \text{ J}$ per unit area of the fault surface.

To estimate the energy required to create fracture surfaces during a single earthquake, we divide the total energy by 10,000 large earthquakes (to account for 44 km of total displacement), which yields $7 \times 10^4 \text{ J m}^{-2}$ per earthquake. This estimate is a lower bound because it does not include the energy associated with refracturing healed grain boundaries and comminuted grains enlarged by Ostwald

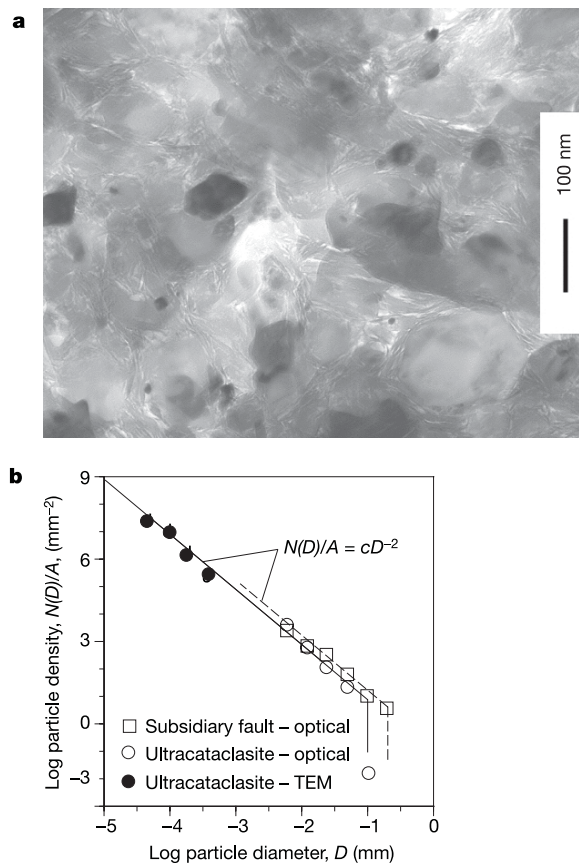


Figure 3 | Particle size of ultracataclasite and subsidiary fault cataclasite. **a**, TEM image of ultracataclasite showing particles 10 to 200 nm in diameter. **b**, Measurements of particle density as a function of particle diameter, where density is given as number of particles in a class size, $N(D)$, divided by area, A . Class sizes from 50–400 nm are imaged using TEM²⁵ and from 6–400 μm are imaged optically¹². All measurements are consistent with a power-law relation with slope of -2 . Solid and dashed lines show particle size distributions assumed for the fracture surface area calculations of ultracataclasite and subsidiary fault cataclasite, respectively.

ripening. As an example, if 50% of the ultracataclasite in the millimetres-thick principal slip surface was refractured during an earthquake, and the energy to refracture a grain boundary is half of that required to fracture an intact grain, then fracture surface energy for a single earthquake increases to approximately $5 \times 10^5 \text{ J m}^{-2}$.

Our estimate of the energy required to create fracture surfaces during a single earthquake is small ($<1\%$) relative to estimates of E_H/S for the Punchbowl fault at a depth of 2–4 km, which even for a coefficient of friction of 0.2 is about $5 \times 10^7 \text{ J m}^{-2}$. This is in direct contrast to a recent conclusion that the surface energy of gouge constitutes half or more of the earthquake energy budget²⁶. Our estimate also cannot account for seismologic values of G (about 10^6 J m^{-2} ; refs 6–8) unless we assume that a significant number of healed grain boundaries and ripened grains are refractured in each successive earthquake, and that all fracturing is attributed to the breakdown phase of rupture. Because some proportion of fracture surface area must be produced during sliding at residual friction²¹, additional processes such as lubrication, flash heating, and thermal pressurization probably contribute to the breakdown in strength^{8–10,23}. These latter processes are compatible with the microstructure of the Punchbowl fault ultracataclasite^{12,25,29}, but would not be recorded by fracture surface area.

The total cumulative fracture surface energy of the Punchbowl fault is about 300 times greater than the fracture energy, G , for a single large earthquake. If G is attributed entirely to the creation of fracture surface, then the observed damage could be produced by earthquake slip totalling less than 1 km of displacement. The differ-

ent estimates of fracture surface area created during seismic slip on a mature fault versus that associated with mining-induced ruptures^{19,20,26} may be explained by extensive fracturing during fault formation and less fracturing associated with slip on an established fault surface. This explanation is consistent with large values of G determined in the laboratory for shear fracture relative to that for frictional sliding^{4,5}. Our results are also consistent with previous conclusions that the overall structure of large-displacement faults is established early in faulting history^{12,30}.

Received 9 January; accepted 15 June 2005.

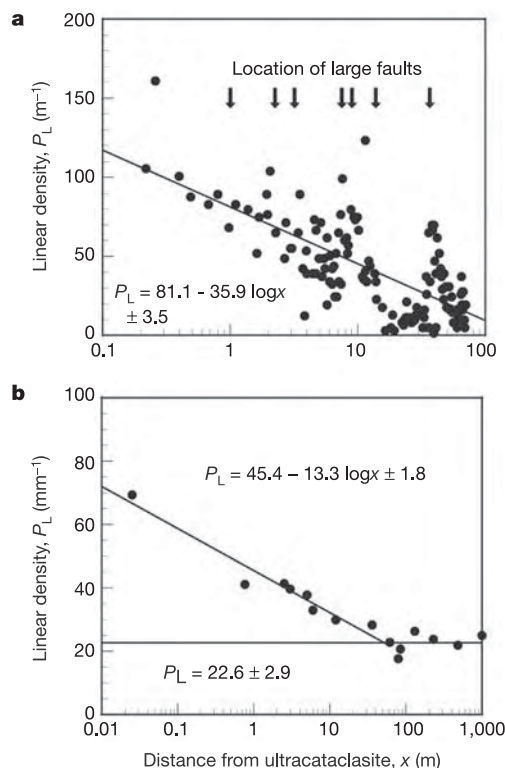


Figure 4 | Subsidiary fault and microfracture density in the damaged Punchbowl Formation along the fault²⁴. Solid line shows best-fit relations (with 95% confidence interval) assumed for calculating fracture surface area. **a**, Linear density of subsidiary faults measured along four traverses across the damage zone (solid circles). Locations of subsidiary faults of the largest size class (gouge layers 1–2 cm thick) are identified with arrows. **b**, Linear density of microfractures (solid circles). At distances greater than approximately 100 m, microfracture density reflects regional density. The difference between near-fault and regional distributions is used to calculate the fracture surface area associated with faulting.

1. Ida, Y. Cohesive force across the tip of a longitudinal-shear crack and Griffith's specific surface energy. *J. Geophys. Res.* **77**, 3796–3805 (1972).
2. Hussein, M. I. & Randall, M. J. Rupture velocity and radiation efficiency. *Bull. Seismol. Soc. Am.* **66**, 1173–1187 (1976).
3. Andrews, D. J. Rupture dynamics with energy loss outside the slip zone. *J. Geophys. Res.* **110**, B01307 (2005).
4. Wong, T. F. Shear fracture energy of Westerly granite from post-failure behaviour. *J. Geophys. Res.* **87**, 990–1000 (1982).
5. Okubo, P. G. & Dieterich, J. H. Effects of physical fault properties on frictional instabilities produced on simulated faults. *J. Geophys. Res.* **89**, 5817–5827 (1984).
6. Guatteri, M., Spudich, P. & Beroza, G. C. Inferring rate and state friction parameters from a rupture model of the 1995 Hyogo-ken Nanbu (Kobe) Japan earthquake. *J. Geophys. Res.* **106**, 26511–26521 (2001).
7. Rice, J. R., Sammis, C. G. & Parsons, R. Off-fault secondary failure induced by a dynamic slip-pulse. *Bull. Seismol. Soc. Am.* **95**, 109–134 (2005).
8. Abercrombie, R. E. & Rice, J. R. Can observations of earthquake scaling constrain slip weakening? *Geophys. J. Int.* **162**, 406 (2005).
9. Heaton, T. H. Evidence for and implications of self-healing pulses of slip in earthquake rupture. *Phys. Earth Planet. Inter.* **64**, 1–20 (1990).
10. Kanamori, H. & Heaton, T. H. in *GeoComplexity and the Physics of Earthquakes* (eds Rundle, J. B., Turcotte, D. L. & Klein, W.) (American Geophysical Union, Washington DC, 2000).
11. Chester, F. M. & Logan, J. M. Composite planar fabric of gouge from the Punchbowl fault, California. *J. Struct. Geol.* **9**, 621–634 (1987).
12. Chester, F. M., Chester, J. S., Kirschner, D. L., Schulz, S. E. & Evans, J. P. in *Rheology and Deformation in the Lithosphere at Continental Margins* (eds Karner, G. D., Taylor, B., Driscoll, N. W. & Kohlstedt, D. L.) (Columbia Univ. Press, New York, 2004).
13. Kanamori, H. Mechanics of earthquakes. *Annu. Rev. Earth Planet. Sci.* **22**, 207–237 (1994).
14. Venkataraman, A. & Kanamori, H. Observational constraints on the fracture energy of subduction zone earthquakes. *J. Geophys. Res.* **109**, B05302 (2004).
15. Beeler, N. M., Wong, T. F. & Hickman, S. H. On the expected relationships among apparent stress, static stress drop, effective shear fracture energy, and efficiency. *Bull. Seismol. Soc. Am.* **93**, 1381–1389 (2003).
16. Brace, W. F. & Walsh, J. B. Some direct measurements of the surface energy of quartz and orthoclase. *Am. Mineral.* **47**, 1111–1122 (1962).
17. Friedman, M., Handin, J. & Alani, G. Fracture-surface energy of rocks. *Int. J. Rock Mech. Mining Sci.* **9**, 757–766 (1972).
18. Wong, T. F. in *Earthquake Source Mechanics* (eds Das, S., Boatwright, J. & Scholz, C. H.) 1–12 (American Geophysical Union, Washington DC, 1986).
19. Olgaard, D. L. & Brace, W. F. The microstructure of gouge from a mining-induced seismic shear zone. *Int. J. Rock Mech. Mining Sci.* **20**, 11–19 (1983).
20. McGarr, A., Spottiswoode, S. M., Gay, N. C. & Ortlepp, W. D. Observations relevant to seismic driving stress, stress drop, and efficiency. *J. Geophys. Res.* **84**, 2251–2261 (1979).
21. Yund, R. A., Blanpied, M. L., Tullis, T. E. & Weeks, J. D. Amorphous material in high strain experimental fault gouges. *J. Geophys. Res.* **95**, 15589–15602 (1990).
22. Scholz, C. H., Dawers, N. H., Yu, J. Z., Anders, M. H. & Cowie, P. A. Fault growth and fault scaling laws: Preliminary results. *J. Geophys. Res.* **98**, 21951–21961 (1993).
23. Andrews, D. J. A fault constitutive relation accounting for thermal pressurization of pore fluid. *J. Geophys. Res.* **107**, ESE15–1–8, art. no. 2363 (2002).
24. Wilson, J. E., Chester, J. S. & Chester, F. M. Microfracture analysis of fault growth and wear processes, Punchbowl Fault, San Andreas System, California. *J. Struct. Geol.* **25**, 1855–1873 (2003).
25. Chester, J. S., Kronenberg, A. K., Chester, F. M. & Guilleminette, R. N. Characterization of natural slip surfaces relevant to earthquake mechanics. *Eos* **84**(46) (2003).
26. Wilson, B., Dewers, T., Reches, Z. & Brune, J. Particle size and energetics of gouge from earthquake rupture zones. *Nature* **434**, 749–752 (2005).
27. Bonnet, E. et al. Scaling of fracture systems in geological media. *Rev. Geophys.* **39**, 347–383 (2001).

28. Underwood, E. E. *Quantitative Stereology* (Addison Wesley, Reading, Massachusetts, 1970).
29. Sibson, R. H. Thickness of the seismic slip zone. *Bull. Seismol. Soc. Am.* **93**, 1169–1178 (2003).
30. Evans, J. P., Shipton, Z. K., Pachell, M. A., Lim, S. J. & Robeson, K. in *Proc. 3rd Conf. on Tectonic Problems of the San Andreas Fault System* (eds Bokelmann, G. & Kovach, R. L.) 67–81 (Stanford Univ. Publ., 2000).

Acknowledgements We thank P. Spudich, T. Heaton and N. Beeler for discussions and review of an early version of this paper, and J. Jenson for advice regarding data analysis. The TEM work was performed in the Microscopy and Imaging Center of Texas A&M University and Z. Luo is acknowledged for his assistance. This research was supported by the Southern California Earthquake Center through a NSF and USGS Cooperative Agreement (J.S.C.), US National Science Foundation (J.S.C. and F.M.C.), and US Geological Survey (J.S.C.).

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.S.C. (chesterj@geo.tamu.edu).

The axial skeleton of the Devonian tetrapod *Ichthyostega*

Per Erik Ahlberg¹, Jennifer A. Clack² & Henning Blom^{1,2}

Ichthyostega was the first Devonian tetrapod to be subject to a whole-body reconstruction^{1–3}. It remains, together with *Acanthostega*⁴, one of only two Devonian tetrapods for which near-complete postcranial material is available. It is thus crucially important for our understanding of the earliest stages of tetrapod evolution and terrestrialization. Here we show a new reconstruction of *Ichthyostega* based on extensive re-examination of original material and augmented by recently collected specimens. Our reconstruction differs substantially from those previously published and reveals hitherto unrecognized regionalization in the vertebral column. *Ichthyostega* is the earliest vertebrate to show obvious adaptations for non-swimming locomotion. Uniquely among early tetrapods, the presacral vertebral column shows pronounced regionalization of neural arch morphology, suggesting that it was adapted for dorsoventral rather than lateral flexion.

A complete skeletal reconstruction of *Ichthyostega* was first presented in 1955 (ref. 2) and republished with a more detailed commentary in 1980 (ref. 3). Jarvik's final reconstruction in 1996 (ref. 5), differs from the previous version only in a few details such as the relatively smaller shoulder girdle⁵. Both reconstructions show an undifferentiated presacral vertebral column, with similar posteriorly

inclined neural arches from the occiput to the pelvis, and posterior to the pelvis a very gradual transition into the more slender but otherwise comparable neural arches of the tail (Fig. 1b). However, the 1996 monograph⁵ contains evidence of problems with this interpretation: the detailed reconstruction of two vertebrae in Jarvik's Fig. 34 shows a neural arch morphology that is not represented anywhere in the complete postcranial reconstruction, and the same applies to the specimen photographs in plates 34:1 and 34:2. Recent work has also cast doubts on the relative proportions of the parts of the skeleton⁶, and the paddle like hindlimb with its unique pattern of seven digits has been described previously⁷.

We have collated all available specimens of the vertebral column of *Ichthyostega*, in order to arrive at a longitudinally near-complete reconstruction (Fig. 1a). Other than a short gap in the proximal part of the tail—possibly spanned by a series of haemal arches—the reconstructed column is complete from occiput to tip of tail. No single specimen shows a complete column. However, the neck, trunk and immediate postsacral region are represented by a series of partial specimens, the preserved regions of which overlap substantially along the body axis (Fig. 2; see Methods and Supplementary Information). The correct relative sizes of the limb girdles and head can also be inferred from this material. All ribs are represented, complete to the

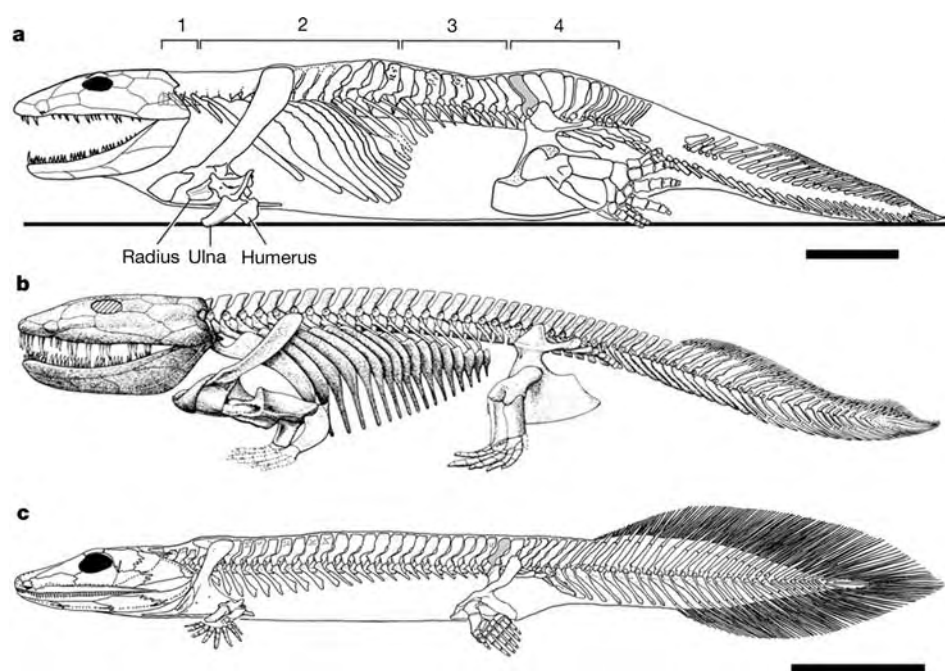


Figure 1 | Reconstructions of *Ichthyostega* and *Acanthostega*. **a**, New reconstruction of *Ichthyostega*, most anterior cervical vertebrae (obscured by skull in lateral view) in grey outline. The reconstruction shows a maximally extended presacral column. Note the revised dentition, orientation of the limb elements and proportional differences (larger limb girdles and shorter tail) compared with **b** (ref. 5). The forearm is shown close to the beginning of the powerstroke; the manus remains unknown. Neural spines of the cervical and anterior caudal regions are also unknown. Regions are as follows: cervical (1), thoracic (2), lumbar (3) and sacral (4); unlabelled, caudal. Hindlimb based on ref. 7 and new observations. **b**, Previous reconstruction of *Ichthyostega* from ref. 5. **c**, Lateral view reconstruction of *Acanthostega* based on unpublished illustrations by Coates and refs 4 and 16. Grey shading in **a** and **c** shows what are considered to be the sacral neural spines. Scale bars, 100 mm.

¹Subdepartment of Evolutionary Organismal Biology, Department of Physiology and Developmental Biology, Uppsala University, Norbyvägen 18A, 752 36 Uppsala, Sweden.

²Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK.

tips apart from two or three in the mid-trunk region. Neural arches are known and essentially complete from the middle of the ribcage to the tip of the tail, except for the previously mentioned gap in the tail base. The intercentra are less well understood, as they are often concealed by other elements, but examples can be seen at intervals throughout the column and these do not appear to differ greatly from the published descriptions⁵. We have not observed any indisputable pleurocentra, with the possible exception of some poorly ossified elements in MGUH (Geological Museum, Copenhagen) VP (Vertebrate Palaeontology) 6140 (Fig. 2). The shape of the neural arch bases provides room for a series of rounded pleurocentra as reconstructed by Jarvik⁵, but it seems that these elements were in fact either unossified or absent. The most important gaps in our data are the incomplete tail base and an almost total lack of information about the cervical neural arches.

The vertebral column differs radically from Jarvik's interpretation. In complete contrast to previous reconstructions, we find that the neural arches are regionalized into discrete thoracic, lumbar, postsacral and caudal morphologies. Thoracic neural arches are straight and posteriorly inclined; these seem to be the model for the uniform presacral neural arches reconstructed by Jarvik⁵, but they are taller

and have rounded rather than square tips. Jarvik's reconstructed morphology appears to be based directly on MGUH VP 6115, in which the tips of the thoracic neural arches are all broken off at a single horizontal plane. Posteriorly the thoracic arches grade rapidly into the upright, anteriorly concave arches of the lumbar region, and become progressively anteriorly inclined nearer the pelvis. The neural arch of the sacral vertebra has the same procumbent morphology, but the first four postsacrals have broad fan-shaped arches that effect a rapid transition in orientation from anteriorly to posteriorly inclined (Fig. 3). The latter morphology, slender and anteriorly concave in contrast to the thoracic neural arches, grades smoothly into the caudal arch morphology. There is also variation in neural spine height, with the thoracic and postsacral vertebrae being significantly taller than the posterior lumbar vertebrae. The different specimens agree well in general structure and show little individual variation. The only significant exception is the lumbar column fragment MGUH VP 6140 (Fig. 2c, d), that seems to lack the swift transition from recumbent to procumbent neural arch orientation seen in, for example, MGUH VP 6054 (Fig. 3). However, MGUH VP 6140 is too short to determine whether this really reflects a major difference in morphology. The presence of broad flanged ribs in the specimen demonstrates that it represents *Ichthyostega* rather than some other taxon.

In addition to the differences in neural arch shape, there is strong regional variation in the development of muscle attachments on the arches. The thoracic, postsacral and caudal arches have no visible muscle scars, but some of the anterior lumbar arches carry large jagged outgrowths that must represent ossified attachments for the myosepta of the epaxial musculature. These can be seen in several

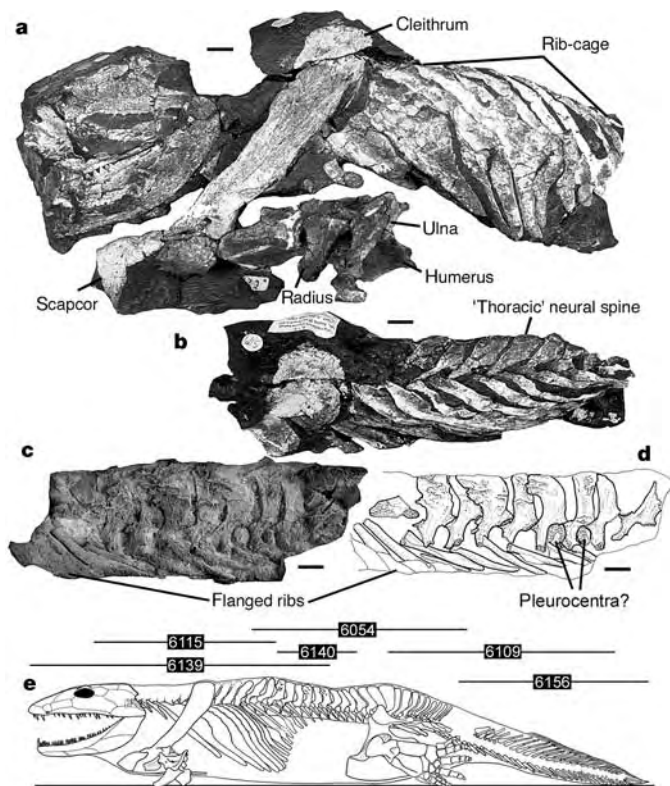


Figure 2 | Specimens of *Ichthyostega*. **a**, MGUH VP 6115 in left lateral view with forearm articulated as close to life position as practical (humerus and ulna are separate elements, radius is impacted onto the ventral surface of the humerus, attitude is slightly elevated from natural position). Note the relative size of the shoulder girdle corresponds with that in Fig. 1a but not Fig. 1b. **b**, Thoracic region of above specimen viewed in right dorsolateral aspect to show backward sloping neural spines. **c**, MGUH VP 6140, isolated length of presacral column, whitened latex peel figured in reverse, anterior to the left, showing alternating pattern of narrow and broad neural spines, the latter with muscle scars and processes overlapping onto next most anterior spine. Thoracic region flanged ribs seen anteriorly locate this specimen in the column relative to other specimens. **d**, Interpretive drawing of **c**. Note that the pleurocentra as interpreted by Jarvik have been indicated, though we are not convinced of their identity. **e**, Schematic representation of key MGUH VP specimens used for the reconstruction. Scale bars, 10 mm. The abbreviation scapcor indicates scapulocoracoid.

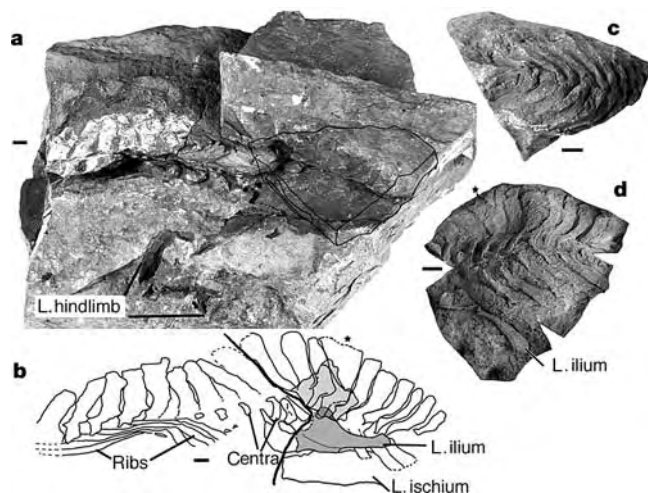


Figure 3 | MGUH VP 6054, vertebral column, pelvis and hindlimb of *Ichthyostega*. **a**, Specimen, in left dorsolateral view, partially assembled showing presacral column as bone, and pelvis, postsacral neural spines and parts of the hind limb as natural mould. Black outline shows the position of the block figured in **c** and **d**. **b**, Composite drawing compiled from observations of all parts of the specimen, not all of which are figured here or are easily visible in photographs. It shows the neural spines, presacral ribs and parts of the pelvis. The left ilium is indicated by dark shading and the right (concealed) ilium is indicated by light shading. Note that the presacral neural spines are deflected to the right, whereas the postsacral spines are deflected to the left, with the point of dislocation indicated by a heavy line. **c**, Block preserving the postsacral region of neural spines, postsacral ribs and the left postiliac process. **d**, Whitened latex peel of **c**, shown in reverse for ease of comparison with **b**. **c** is a dome-shaped block making the elements appear foreshortened. The peel has been relieved to reduce the three-dimensionality. The starred neural spine in **b** and **d** is the broadest, but in **d** it can clearly be seen to be postsacral. The sacral spine is probably that immediately posterior to the break. Scale bars, 10 mm.

specimens, such as MGUH VP 6139 (Fig. 4) and 6054, but they are best preserved in MGUH VP 6140 (Fig. 2). In this individual the attachments are strongly developed on every other neural arch, and these bony outgrowths extend forward to overlap the next anterior neural arch laterally. We discuss the functional implications of this pattern below. *Acanthostega* also has ossified muscle attachments on some neural spines (Fig. 1c), but these occur throughout the presacral column and do not overlap adjacent vertebrae⁴.

The rib morphology of *Ichthyostega* differs from that reconstructed by Jarvik, although less dramatically than the neural arches. MGUH VP 6139 shows a series of four cervical ribs that lack flanges and become progressively shorter towards the anterior; whereas Jarvik reconstructed no neck at all and showed the longest thoracic rib attaching to the first vertebra (Fig. 1b). The longest of the flanged thoracic ribs is the sixth or seventh (not the first), and the posterior transition into short unflanged lumbar ribs is sharper than in the old reconstructions. There are also fewer flanged ribs in the series than reconstructed by Jarvik. Together, these features give quite a different outline to the ribcage. The thoracic ribs are only gently curved, and the distal ends of the left and right ribs often meet in the midline (for example in MGUH VP 6115 and 6139). This considered along-

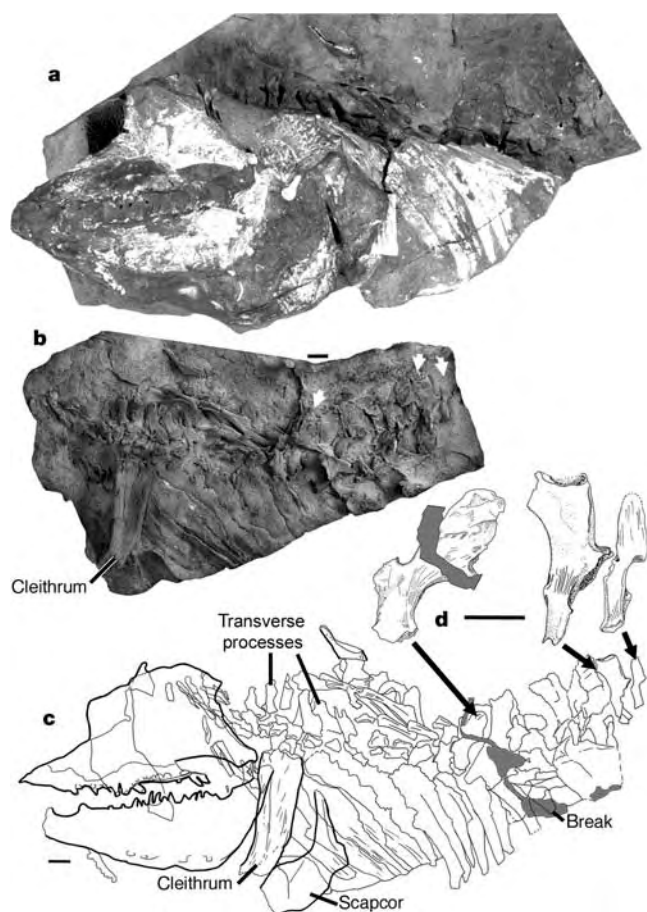


Figure 4 | MGUH VP 6139, skull, shoulder girdle, thoracic and presacral region of *Ichthyostega*. **a**, Specimen part-assembled with skull and thoracic region placed on the natural mould of the right side, in which rib heads and some neural arches are just visible beneath. **b**, Whitened latex peel shown in reverse of basal block shown in **a**. White arrows indicate three neural arches enlarged in **d**. **c**, Composite drawing incorporating information from both right and left sides, showing position of skull, shoulder girdles, rib cage and neural spines. Arrows indicate three neural arches enlarged in **d**. Note the steeply inclined angle of the zygapophyses and the alternation of spine anteroposterior length. **d**, Enlarged neural arches, aligned with ventral part of each arch vertical to highlight differences in neural spine morphology. Scale bars, 10 mm.

side the comparably modest curvature of the cleithrum and scapulocoracoid (for example, Jarvik's plate 45:4, 5; ref. 5), suggests an upright oval body cross-section. The curiously large postsacral ribs were figured accurately by Jarvik.

Our body reconstruction (Fig. 1a) incorporates recent data on the morphology and orientation of the limbs^{6–8} as well as a reassessment of the dentition. Its distinctive morphology differs greatly not only from Jarvik's reconstructions but from other early tetrapods such as *Acanthostega*⁴ (Fig. 1c). *Ichthyostega* presents the earliest example of a differentiated presacral vertebral column in a tetrapod—cervical, thoracic, lumbar, sacral and caudal regions can be distinguished. This morphology further implies a differentiation of the axial musculature unexpected in such a primitive tetrapod, and raises questions about the locomotory mechanics and mode of life of *Ichthyostega*.

We hypothesize that this unique morphology reflects a loss of horizontal flexion in the presacral column and possibly the acquisition of limited vertical flexion in the lumbar region. The deeply overlapping thoracic ribs probably did not permit either lateral or dorsoventral flexion of the thorax. In the lumbar region, the presence of expanded muscle scars that cause some neural arches to overlap adjacent ones (Fig. 2c, d) argues against lateral flexion, and also demonstrates the presence of a strongly developed epaxial musculature. The procumbent orientation of the lumbar neural spines may be an indication of vertical flexion, although the precise relationship between spine orientation and flexion remains unclear even in mammals (some of which have a remarkably *Ichthyostega*-like pattern of neural spine morphologies), and requires further investigation. The zygapophyses of *Ichthyostega* are rarely well preserved, but MGUH VP 6139 (Fig. 4d) shows them to be steeply oriented and certainly not expanded horizontally. We infer that the whole presacral vertebral column was incapable of significant lateral flexion, but that a modest degree of vertical flexion may have occurred in the lumbar region.

If we are correct and the axial morphology did not permit lateral trunk flexion in *Ichthyostega*, two gaits seem theoretically possible: 'walking' with diagonally synchronized limb movements and rigid elevated trunk; and a bilaterally symmetrical 'shuffling' or 'inchworm' movement, with vertical flexion of the lumbar region as part of the powerstroke for the forelimb and recovery stroke for the hindlimb, and lumbar extension contributing to the converse. Both may have been used at different times. These locomotory hypotheses remain to be tested by further analysis of limbs and joint surfaces.

Acanthostega, in contrast to *Ichthyostega*, had a laterally flexible trunk and appears better adapted for tail-propelled swimming. The limbs and limb girdles of *Ichthyostega* are also proportionately much larger than those of *Acanthostega* (Fig. 1a, c). Coupled with differences in the dentition (recurved sectorial teeth in *Ichthyostega*, conical piercing teeth in *Acanthostega*) and body size (*Ichthyostega* is about 30% larger than *Acanthostega*), plus the fact that the two genera rarely occur together in the same locality within the Greenland sediments⁹, this points to a clear ecological separation between these primitive tetrapods.

Despite its peculiarities, which includes an ear specialized for underwater hearing¹⁰, *Ichthyostega* is not completely unique. A ribcage of flanged ribs followed by a more flexible lumbar region is also seen in the Early Carboniferous stem-group tetrapods *Pederpes* and *Whatcheeria*^{11–13}. Recent phylogenetic analyses placing *Ichthyostega* immediately below *Tulerpeton* and these Carboniferous forms¹⁴ suggest that such similarities may be homologies reflecting the gradual terrestrialization of the tetrapod stem group. However, the zygapophyses of *Pederpes* are horizontal¹³ in contrast to the steeply angled orientation in *Ichthyostega*, suggesting a very different mode of locomotion in the two. The variable neural spine heights of *Ichthyostega* (although not the diversified neural spine morphologies) are matched by the Carboniferous crown-group tetrapod *Proterogyrinus*¹⁵ and may relate to terrestrial locomotion. Gradually,

Ichthyostega is emerging from its iconic but isolated position as 'the earliest tetrapod' to occupy a recognizable place in the developing picture of early tetrapod diversity and evolution. *Ichthyostega* appears to be an early and ultimately unsuccessful attempt at adapting the tetrapod body plan for terrestrial locomotion, divergent but not very remote from the lineage that successfully solved these adaptive problems and ultimately gave rise to all living tetrapods.

METHODS

Specimens and techniques used to compile the reconstruction. All specimens carry MGUH prefixed numbers plus either VP or f.n. (field number). Most of the specimens exhibit some degree of vertical compression because of sedimentary compaction (resulting, in morphological terms, in dorsoventral, lateral or oblique compression depending on the orientation of the body within the sediment) but none shows any sign of oblique shear. The observed differences in neural spine orientation relate directly to differences in morphology, and are not accompanied by distortions in adjacent structures such as the ventral parts of the neural arches or the rib heads. We are thus confident that the morphological variation is real and not a product of post-mortem distortion by geological processes.

VP 6054. Block containing the natural mould of a presacral column, pelvic girdle, both hindlimbs, sacral region and first few caudal vertebrae (plate 34 in ref. 5). The specimen was broken and somewhat twisted in the immediately presacral region. In many parts, some etched with dilute HCl and peeled with latex. Sederholm Bjerg, profile 4, 1,174 m, collected in 1934.

VP 6109. Block showing one hindlimb, parts of the pelvis and postsacral column and complete caudal series (plates 37, 38, 65 and 66 in ref. 5). N. Celsius Bjerg, topmost coulisse, collected in 1948.

VP 6138. Block containing obliquely compressed skull, shoulder girdles, both forelimbs and parts of their carpal regions, parts of cervical region (as natural moulds or embedded bone), complete thoracic region and one associated hindlimb. In many parts, with skull, both forelimbs and ribcage partly etched and peeled as above. Sederholm Bjerg, profile 4, 1,174 m, collected in 1949.

VP 6139. Block containing laterally compressed partial skull and shoulder girdle, cervical series (mostly embedded in matrix), thoracic region and part of presacral column (plate 24 in ref. 5). In many parts, with main block etched and peeled as above. Sederholm Bjerg, profile 4, 1,174 m, collected in 1949.

VP 6140. Length of isolated vertebrae that was etched and peeled as above (part figured in plate 34.3 in ref. 5). Sederholm Bjerg, profile 4, 1,174 m, collected in 1949.

VP 6115. Posterior half of skull, both forelimbs and shoulder girdles, somewhat compressed, left with radius and ulna in articulation, thoracic region (plates 42, 45, 46, 51, 53 and 54 in ref. 5). Associated with a hindlimb and partial column that is probably not from the same individual. S. Smith Woodward Bjerg, 370 m, collected in 1948.

VP 6156. Complete tail in part and counterpart (plate 40 in ref. 5). Sederholm Bjerg, profile 4, 1,174 m, collected in 1949.

f.n. 1394. Hindlimb, pelvis and part of postsacral and caudal series. Surface prepared mechanically. Figured in ref 7. Gauss Halvø, 90 m, collected in 1987.

f.n. 300. Two individuals in the same block, one with humerus, shoulder girdles, complete column and ribs up to sacral region, pelvis and hindlimb. Partly surface prepared mechanically. S. Celsius Bjerg, 500 m, collected in 1998.

Note that 6054, 6138, 6139, 6140 and 6156 derive from the same 'in situ' horizon.

Received 24 February; accepted 6 June 2005.

1. Jarvik, E. On the fish-like tail in the ichthyostegid stegocephalians. *Meddelelser om Grønland* **114**, 1–90 (1952).
2. Jarvik, E. The oldest tetrapods and their forerunners. *Sci. Mon.* **80**, 141–154 (1955).
3. Jarvik, E. *Basic Structure and Evolution of Vertebrates* Vol. 1 (Academic, New York, 1980).
4. Coates, M. I. The Devonian tetrapod *Acanthostega gunnari* Jarvik: postcranial anatomy, basal tetrapod relationships and patterns of skeletal evolution. *Trans. R. Soc. Edinb. Earth Sci.* **87**, 363–421 (1996).
5. Jarvik, E. The Devonian tetrapod *Ichthyostega*. *Fossils and Strata* **40**, 1–206 (1996).
6. Coates, M. I. & Clack, J. A. in *Studies on Early Vertebrates* (eds Arsenault, M., Lelièvre, H. & Janvier, P.) 373–388 (Bulletin du Muséum national d'histoire naturelle, Paris, 1995).
7. Coates, M. I. & Clack, J. A. Polydactyly in the earliest known tetrapod limbs. *Nature* **347**, 66–69 (1990).
8. Clack, J. A., Blom, H. & Ahlberg, P. E. New insights into the postcranial skeleton of *Ichthyostega*. *J. Vert. Paleontol. Abstr.* **23**, 41A (2003).
9. Blom, H., Clack, J. A. & Ahlberg, P. E. Localities, distribution and stratigraphical context of the Late Devonian tetrapods of East Greenland. *Meddelelser om Grønland Geosci.* **43**, 4–50 (2005).
10. Clack, J. A. *et al.* A uniquely specialised ear in a very early tetrapod. *Nature* **425**, 65–69 (2003).
11. Clack, J. A. An early tetrapod from Romer's Gap. *Nature* **418**, 72–76 (2002).
12. Lombard, R. E. & Bolt, J. R. A new primitive tetrapod, *Whatcheeria deltae*, from the Lower Carboniferous of Iowa. *Palaeontology* **38**, 471–494 (1995).
13. Clack, J. A. & Finney, S. M. *Pederpes finneyae*, an articulated tetrapod from the Tournaisian of western Scotland. *J. Syst. Palaeontol.* **2**, 311–346 (2005).
14. Ruta, M., Coates, M. I. & Quicke, D. L. J. Early tetrapod relationships revisited. *Biol. Rev.* **78**, 251–345 (2003).
15. Holmes, R. The Carboniferous amphibian *Proterogyrinus scheeli* Romer, and the early evolution of tetrapods. *Phil. Trans. R. Soc. Lond. B* **306**, 431–524 (1984).
16. Clack, J. A. The dermal skull roof of *Acanthostega*, an early tetrapod from the Late Devonian. *Trans. R. Soc. Edinb.* **93**, 17–33 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the Geological Museum and University of Copenhagen for access to their *Ichthyostega* material, and G. Cuny and D. Harper for their help and support during our visits. We would also like to acknowledge the late E. Jarvik and S. Bendix-Almgreen, both of whom gave us access to specimens then in their care. Preparation of MGUH f.n. 300 was carried out by S. Finney. M. Coates kindly permitted us to use one of his unpublished *Acanthostega* reconstructions for part of Fig. 1c. Most of our project was supported by an NERC grant with additional funding for H.B. provided by the Isaac Newton Trust Fund and by a project grant and salary from Vetenskapsrådet (the Swedish Research Council) for P.E.A.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to P.E.A. (per.ahlberg@ebc.uu.se).

Sporadic immunogenic tumours avoid destruction by inducing T-cell tolerance

Gerald Willmsky¹ & Thomas Blankenstein^{1,2}

The recognition and elimination of tumours by T cells, a process termed cancer immunosurveillance¹, is effective against certain virus-associated cancers². Spontaneous tumours often induce a specific immune response and are therefore also immunogenic. However, it is not clear whether they can be controlled by T cells^{3–10}. The immunosurveillance hypothesis postulates that tumours, if they eventually grow, escaped T-cell recognition by losing immunogenicity^{6–8}. Here we show, by generating a mouse model of sporadic cancer based on rare spontaneous activation of a dormant oncogene, that immunogenic tumours do not escape their recognition but induce tolerance. In this model, tumours derive from single cells and express a tumour-specific transplantation rejection antigen. Whereas vaccinated mice remain tumour-free throughout their lifetime, naive mice always develop a progressively growing tumour. We also show that despite specific recognition by T cells, the tumours do not lose their intrinsic immunogenicity and are rejected after transplantation in T-cell-competent recipients. Furthermore, in the primary host tumour-induced tolerance is associated with the expansion of non-functional T cells. Together, our data argue against immunosurveillance of spontaneous cancer.

We developed a sporadic tumour model^{11,12} that better resembles physiological tumour development and should reveal most clearly the ability of the immune system to recognize and eliminate spontaneous tumours. In this model a dormant oncogene, the immunologically well-characterized SV40 T antigen (Tag)^{13,14}, is activated by rare stochastic events in individual cells that express a tumour-specific transplantation rejection antigen, against which the mice should not have developed tolerance. C57BL/6 (B6) transgenic mice were constructed that contained Tag in a silent form due to an attenuator separating a ubiquitously active promoter from the oncogene (Fig. 1a). Additionally, the construct allowed somatic Tag activation by Cre/LoxP-mediated excision of the attenuator (see below). Both lines (LoxP-Tag#9, used for the subsequent experiments, and LoxP-Tag#16) spontaneously developed, without exception, a tumour at a variety of anatomical sites (such as kidney, liver, testis, connective tissue, spleen and thymus). Of a total of 184 mice analysed, in most cases one tumour per mouse was macroscopically identified after a usually long latency, even though in mice with large tumour burden secondary tumours or metastases were occasionally observed. Tumours stained positive with an anti-Tag monoclonal antibody (mAb) on tumour tissue sections (Fig. 1b). Even though the

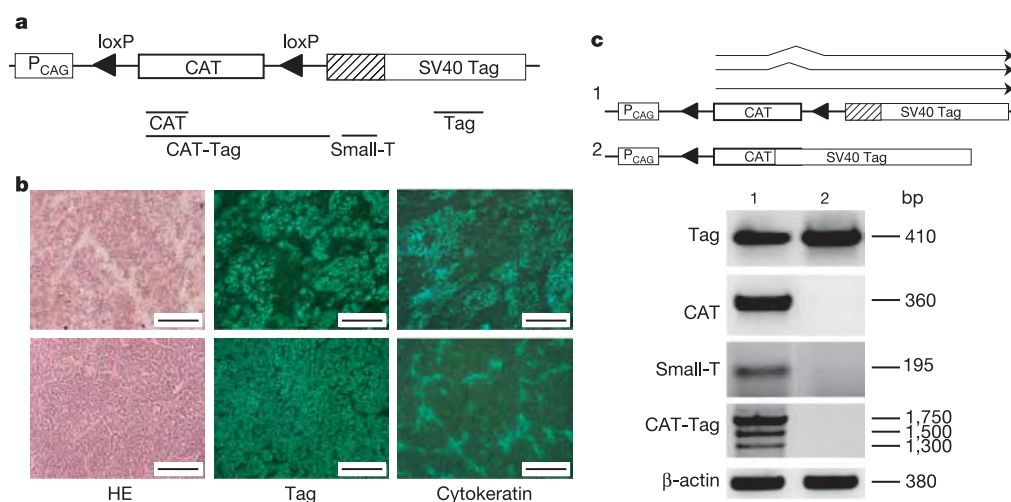


Figure 1 | LoxP-Tag transgenic mice develop sporadic tumours as a result of rare epigenetic or genetic events. **a**, DNA construct used for transgenic mice with a chimaeric β -actin/ β -globin promoter, the CAT gene as attenuator flanked by loxP recognition sites, allowing Cre-mediated excision, and the SV40 T-antigen (Tag). The lines indicate the regions of PCR products amplified in **c**. **b**, Immunohistochemical analysis of spontaneous tumours. Serial cryosections of two representative tumours were stained with haematoxylin/eosin (HE), FITC-labelled Tag-specific antibody (middle) and cytokeratin/FITC anti-mouse IgG1 antibody (right). See also

Supplementary Fig. 1. Scale bars, 100 μ m. **c**, RT-PCR analysis of early-passage tumour cell lines reveals the presence of differentially spliced bicistronic mRNA in some tumours (lane 1) and Cre-recombinase independent deletion of the attenuator region in others (lane 2), determined by partially sequenced PCR products of Tag, CAT, small-T and bicistronic CAT-Tag. Of ten tumours analysed, eight showed bicistronic mRNA and two spontaneous attenuator deletion. β -Actin RT-PCR was used as internal control. bp, base pairs.

¹Institute of Immunology, Charité Campus Benjamin Franklin, Hindenburgdamm 30, 12200 Berlin, Germany. ²Max Delbrück Center for Molecular Medicine, Robert-Roessle Strasse 10, 13122 Berlin, Germany.

cellular origin of the tumours is not entirely clear, some tumours stained positive for cytokeratin, an epithelial and mesothelial marker, whereas others were CD3⁺, B220⁺, CD11b⁺, indicating a myeloid origin (Fig. 1b and Supplementary Fig. 1). In some tumours, Tag mRNA was detected on a bicistronic message as a result of a read-through mechanism (Fig. 1c). In others, spontaneous—Cre/LoxP independent—deletion of the attenuator including the small-T region was detected, probably placing Tag in closer proximity to the promoter. Tumour cells with bicistronic mRNA expressed CAT, the attenuator gene, whereas tumours with spontaneous attenuator deletion did not express CAT (Supplementary Fig. 2). LoxP-Tag mice therefore develop, by genetic and/or epigenetic mechanisms, a sporadic tumour quasi-randomly in a variety of different organs

and, at least to some extent, these tumours are of different cellular origin.

Because we could not exclude the possibility that leakiness of the attenuator resulted in Tag expression in addition to that found in the tumour that could result in tolerance, we analysed the Tag-specific immune competence in young tumour-free transgenic mice. LoxP-Tag and B6 mice were immunized with interleukin (IL)-7/B7.1/Tag transfected MCA-205 fibrosarcoma cells and challenged with Tag-expressing B16.F10 melanoma cells. In contrast to naive mice, most immunized B6 and LoxP-Tag mice rejected B16.F10-Tag⁺ but not B16.F10-mock tumour cells, demonstrating Tag-specific rejection (Fig. 2a). Additionally, T cells specific for the dominant Tag epitope peptide IV were analysed with H-2K^b/IV

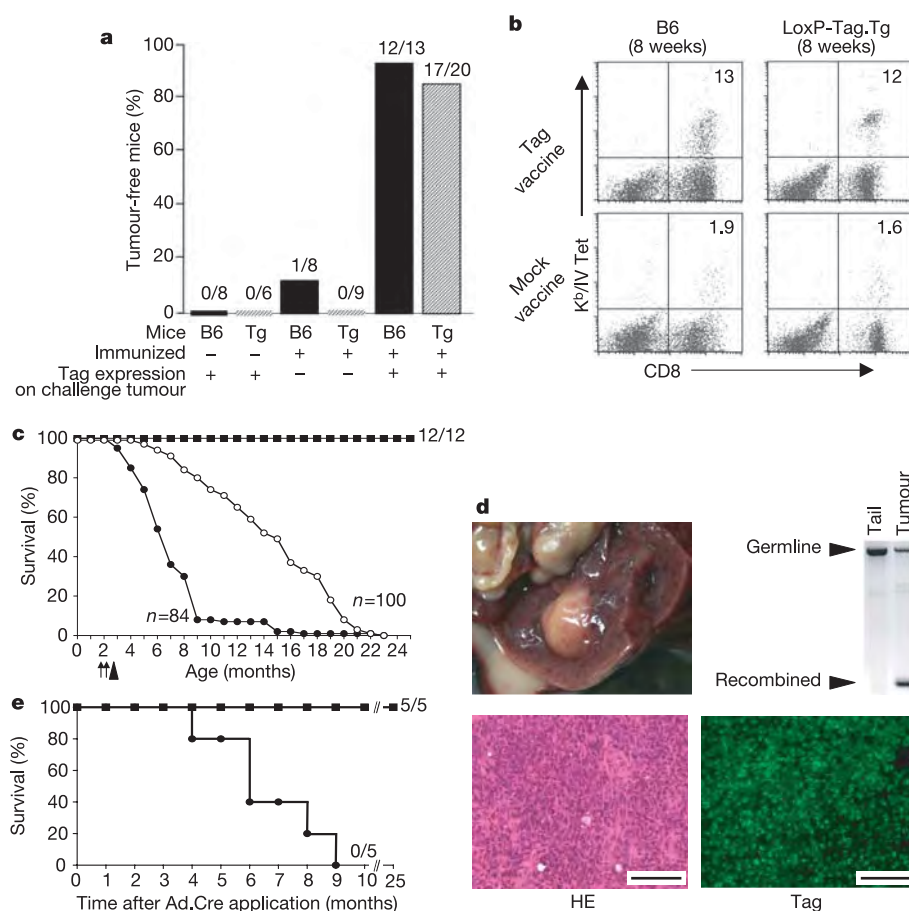


Figure 2 | LoxP-Tag mice are not tolerant towards Tag and control sporadic and Cre-recombinase-induced tumours after immunization. **a**, Rejection of transplanted Tag-positive tumours. As indicated, 8-week-old wild-type (B6, black bars) and LoxP-Tag transgenic (Tg, hatched bars) mice were immunized twice with 2.5×10^5 IL-7/B7.1/Tag transfected irradiated MCA-205 cells or left untreated and challenged two weeks later contralaterally with 2.5×10^5 Tag-expressing or mock-transfected B16.F10 melanoma cells and tumour growth was monitored. Shown are the percentages of tumour-free mice at the end of experiments. The number of tumour-free mice per number of mice from two combined experiments is indicated. **b**, Expansion of Tag epitope IV-specific CD8⁺T cells after Tag-specific immunization. Indicated mice were immunized once as above. After two weeks, spleen cells were stimulated *in vitro* for 5–7 days with Tag-specific peptide IV and stained with anti-CD8 mAb and H2-K^b/epitope IV tetramers. The percentage of CD8⁺T cells that were H2-K^b/epitope IV-positive are indicated. One of three experiments with similar results is shown. **c**, Immunized LoxP-Tag mice shown in **a** (squares; $n = 12$) and naive LoxP-Tag mice were monitored for spontaneous tumour development. The survival of mice out of the number of immunized mice from two combined experiments is indicated. Immunized mice were macroscopically confirmed

to be tumour-free at the end of experiment. The arrows and arrowhead indicate immunization and B16-Tag⁺ challenge, respectively. For control LoxP-Tag mice two plots are shown: the first-generation progenies succumbed to a tumour with an average latency of 6 months (filled circles; $n = 84$; between 1999 and 2001); the following generations, for an unknown reason, did so with an average latency of about 14 months (open circles; $n = 100$; between 2001 and 2003). These mice were confirmed to have developed a macroscopically visible tumour when they were moribund. **d**, Tag activation and induction of liver tumours by adenovirus-mediated Cre-recombination. Eight-week-old LoxP-Tag transgenic mice were injected intravenously with 10^9 plaque-forming units of Ad.Cre. When mice became moribund, tumours in the liver were macroscopically identified (upper left), stained with haematoxylin/eosin (lower left) and FITC-labelled Tag-specific antibody (lower right). Scale bars, 100 μ m. For analysis of Cre-mediated loxP-recombination, genomic DNA from tail biopsies and liver tumour tissue were subjected to recombination-specific PCR with 5' promoter and 3' Tag-specific primers (upper right). Tag expression and attenuator deletion after Ad.Cre injection were found in five of five tumours. **e**, Each of five naive (circles) and immunized (squares) (see above) LoxP-Tag transgenic mice was injected intravenously with Ad.Cre and monitored for tumour growth.

tetramers¹⁴. In Tag-vaccinated B6 and LoxP-Tag mice the frequency of K^b/IV-tetramer⁺ cells increased to $13 \pm 5\%$ and $12 \pm 5\%$ (s.d.) of the CD8⁺T cells, respectively, in comparison with 1.9% and 1.6% in mock-vaccinated mice (Fig. 2b). Thus, leakiness of the attenuator that inhibits Tag expression is too rare a phenomenon or of too low abundance to induce tolerance.

Next we analysed whether LoxP-Tag mice could control the autochthonous tumour if they were immunized before tumour development. Tag-immunized mice remained tumour-free (12 of 12) throughout the observation period of two years (Fig. 2c). Non-immunized LoxP-Tag control mice all had to be killed because of overt signs of a tumour before this time point. This demonstrates that LoxP-Tag mice efficiently controlled primary tumours if they were immunized in a timely manner. The failure of T cells to control primary tumours is therefore not due to their inability to do so. To confirm these results, immunized LoxP-Tag mice were injected with a Cre-recombinase-encoding adenovirus (Ad.Cre) that effectively deleted the attenuator and activated Tag. All naive LoxP-Tag mice injected with Ad.Cre developed, as a result of adenovirus tropism, liver tumours that showed deletion of the attenuator because of Cre/LoxP-mediated recombination and Tag⁺ staining on tissue sections (Fig. 2d). Ad.Cre-treated immunized LoxP-Tag mice remained tumour-free during 25 months' observation (Fig. 2e), indicating that after immunization they could efficiently control the large numbers of Tag⁺ cells that would form progressively growing tumours.

We investigated whether and when the primary tumour induced a

specific immune response. A time-course analysis of serum antibodies in LoxP-Tag mice revealed that they developed Tag-specific IgG antibodies that increased over time to very high levels (up to 3 mg ml^{-1}) until the mice became moribund (Fig. 3a). These antibodies seemed to be a prognostic marker for progressive disease; they illustrate that the tumour was recognized several months before the mice died. Anti-Tag antibody concentrations in tumour-bearing LoxP-Tag mice were even higher than in young tumour-free mice after Tag-specific immunization (Fig. 3b). Tumour-bearing double-transgenic LoxP-Tag $\times$ Alb-Cre mice, which express the Cre-recombinase by the albumin promoter and are tolerant for Tag, did not produce Tag-specific antibodies. Importantly, K^b/IV-tetramer⁺ CD8⁺T cells spontaneously expanded in tumour-bearing LoxP-Tag mice in a comparable manner to Tag-immunized tumour-free LoxP-Tag mice ($10 \pm 6\%$) (Fig. 3c). Tumour-infiltrating CD4⁺ and CD8⁺T cells were readily detected by immunohistology (Supplementary Fig. 3). It therefore seems that the sporadic tumour induces a strong and persistent humoral and cellular immune response.

The progressive growth of tumours in the presence of a tumour-specific immune response raised the question of whether the tumours evaded their destruction by becoming poorly immunogenic. We therefore established 13 tumour-cell lines from LoxP-Tag mice that all expressed Tag (data not shown). Of the 13 cell lines, 11 grew progressively as a tumour in Rag-1^{-/-} or severe combined immunodeficiency (SCID) mice (Supplementary Table 1). All of these 11 tumours were rejected in naive B6 mice, and 4 of 4 tumours

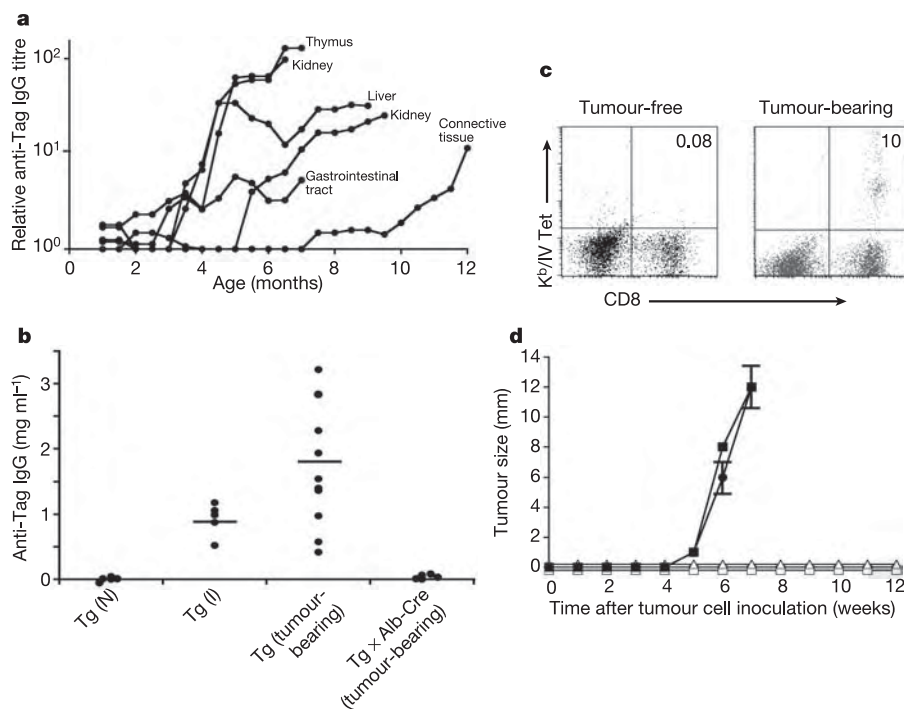


Figure 3 | Sporadic immunogenic LoxP-Tag tumours induce a Tag-specific immune response. **a**, Serum was obtained from LoxP-Tag transgenic mice every two weeks and relative Tag-specific IgG antibody titres, shown on a logarithmic scale, were determined by ELISA. When mice became moribund, the site of tumour growth was identified macroscopically. **b**, The end-point concentration of Tag-specific IgG antibodies was determined from serum obtained from individual mice (mean values are indicated by bars). Shown are naive (N; $n = 5$) and immunized (I; $n = 5$) young (6–8 weeks old) LoxP-Tag transgenic mice (Tg), tumour-bearing LoxP-Tag transgenic mice (10–20 months old; $n = 10$) and tumour-bearing LoxP-Tag $\times$ Alb-Cre double-transgenic mice (6–12 weeks old; $n = 5$). **c**, Spleen cells of tumour-free (8 weeks old; left) and tumour-bearing (31 weeks old;

right) LoxP-Tag transgenic mice were stimulated *in vitro* for 5–7 days with Tag-specific peptide IV and stained with anti-CD8 mAb and H2-K^b/epitope IV tetramers. The percentages of CD8⁺T cells that were positive for H2-K^b/epitope IV are indicated. One of three experiments that showed between 3% and 18% tetramer/CD8-positive cells in tumour-bearing LoxP-Tag mice is shown. **d**, Sporadic tumours of LoxP-Tag mice are highly immunogenic because of Tag expression. Cells (10^6) of tumour line 9.79 (and 16.113; shown in Supplementary Fig. 4) were subcutaneously injected into 8–12-week-old SCID mice (filled squares; $n = 3$), C57Bl/6 mice (triangles; $n = 3$), LoxP-Tag mice (open squares; $n = 2$) and 8-week-old Tag-tolerant LoxP-Tag $\times$ Alb-Cre mice (circles; $n = 3$) and tumour growth was observed. Error bars represent s.d.

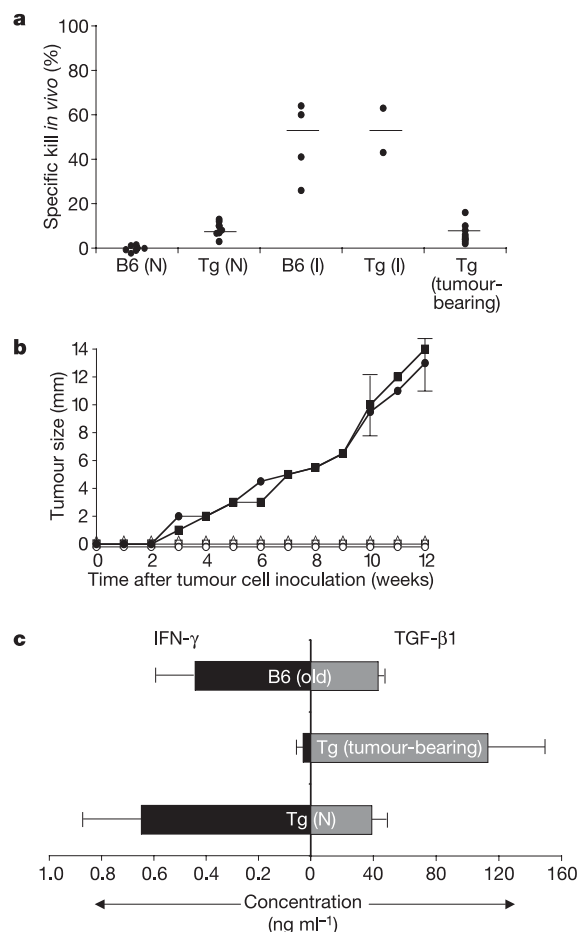


Figure 4 | Sporadic immunogenic tumours induce T-cell tolerance. **a**, CTL activity against the Tag-specific peptide IV was analysed *in vivo*. Non-loaded and peptide-loaded spleen cells (10^7 each) labelled with different amounts of CFSE were injected into the indicated mice and 18 h later the ratio between both populations was determined by flow cytometry. The percentage of specific killing of peptide-loaded cells is indicated. Shown are the combined data of all experiments (bars indicate mean values), and one representative example per experimental group is shown in Supplementary Fig. 5. Naive 8–12-week-old C57Bl/6 mice (B6 (N); $n = 7$), naive 8–12-week-old LoxP-Tag mice (Tg (N); $n = 6$), immunized 14–18-week-old C57Bl/6 mice (B6 (I); $n = 4$), immunized 20-week-old LoxP-Tag mice (Tg (I); $n = 2$) and 14–18-month-old tumour-bearing LoxP-Tag transgenic mice ($n = 9$) were analysed. Immunization was performed by single subcutaneous injection of 10^6 cells of tumour line 16.113. **b**, Young tumour-free but not old tumour-bearing LoxP-Tag mice reject Tag⁺ immunogenic tumours. Cells (10^6) of tumour line 16.113 were injected subcutaneously into SCID mice (filled squares; $n = 2$), young C57Bl/6 mice (triangles; $n = 5$), old C57Bl/6 mice (open circles; $n = 10$), young LoxP-Tag mice (open squares; $n = 5$) and old LoxP-Tag mice (filled circles; $n = 7$) and tumour growth was monitored. Young mice were 3-month old; old mice were 7–16-month old (LoxP-Tag) or 14-month old (C57Bl/6). In six of the seven old LoxP-Tag mice the transplanted tumour grew (error bars represent s.d.). Tag-specific antibodies before tumour challenge and growth of challenge and spontaneous tumour are shown in Supplementary Fig. 6. **c**, Tumour-bearing LoxP-Tag mice show severely reduced serum levels of IFN- γ but increased serum levels TGF- β 1. Young LoxP-Tag mice (3 months old; Tg(N)), tumour-bearing LoxP-Tag mice (14–22 months old) and old B6 mice (18 months old) (three per group) were analysed by an *in vivo* IFN- γ capture assay for serum IFN- γ levels. In addition, serum obtained from individual young LoxP-Tag mice (3 months old; $n = 3$), tumour-bearing LoxP-Tag transgenic mice (14–22 months old; $n = 8$) and old B6 mice (15 months old; $n = 3$) was analysed for TGF- β 1 by ELISA. Error bars represent s.d.

were rejected in naive young LoxP-Tag mice, showing the high immunogenicity of the tumour lines. To analyse whether the immunogenicity was due to Tag expression, two tumour cell lines were injected into Tag-tolerant LoxP-Tag $\times$ Alb-Cre transgenic mice. As shown in Fig. 3d and Supplementary Fig. 4, tumours grew in LoxP-Tag $\times$ Alb-Cre mice and in SCID mice (as control) but not in B6 or young LoxP-Tag mice, indicating that Tag might have been the dominant rejection antigen responsible for the high immunogenicity. These results confirm that LoxP-Tag mice have retained normal immune competence against Tag. They also show that highly immunogenic tumours do not have to escape immune recognition during progression in their primary immunocompetent host.

The progressive growth of highly immunogenic tumours in the presence of a specific immune response raised the question of whether Tag-specific T cells in tumour-bearing LoxP-Tag mice were functional. We therefore used an *in vivo* kill assay. Tag-peptide IV loaded and non-loaded control spleen cells, labelled with different amounts of 5(6)-carboxyfluorescein diacetate, succinimidyl ester (CFSE), were injected into tumour-bearing LoxP-Tag mice and, as control, naive or Tag-immunized B6 and young tumour-free LoxP-Tag mice; the disappearance of CFSE-labelled cells was analysed. Tag-peptide loaded cells decreased significantly in immunized mice, but not in naive B6 and young LoxP-Tag mice, in comparison with CFSE-labelled control spleen cells (Fig. 4a and Supplementary Fig. 5). Tumour-bearing LoxP-Tag mice with substantially expanded Tag-specific CD8⁺T cells (see above) did not kill peptide-loaded spleen cells at more than the background rate, indicating that they were non-functional. To confirm these data, the immunogenic tumour line 16.113 was injected into old LoxP-Tag mice (Fig. 4b), most of which had developed already anti-Tag antibodies as an indication of a sporadic tumour (Supplementary Fig. 6). As before, tumours grew in SCID mice but were rejected in B6 or young LoxP-Tag mice. In six of seven old LoxP-Tag mice, 16.113 tumours grew, showing that they had lost the ability to reject immunogenic tumours and had acquired systemic tolerance. It is not yet known whether the primary tumour permits the growth of the transplanted tumour in a Tag-specific manner¹⁵. Because old B6 mice still rejected the tumour (Fig. 4b), the failure of old LoxP-Tag mice to reject immunogenic Tag⁺ tumours was tumour-induced and was not due to a general age-dependent decline of immune competence.

To improve our understanding of why Tag-specific B cell responses were undiminished but T cells seemed to be anergic, the anti-Tag antibody subtypes in the serum of tumour-bearing LoxP-Tag mice were informative. In almost all mice, Tag-specific IgG2a, IgG2b and occasionally IgA, but no IgG1 or IgG3, were detected (Supplementary Fig. 7). IgG2a indicates interferon (IFN)- γ involvement in the anti-tumour response, and IgG2b and IgA indicate the involvement of transforming growth factor (TGF)- β 1. Paradoxically, both cytokines inhibit each other reciprocally, and TGF- β 1 effectively suppresses T-cell function and induces local inflammation^{16,17}. To analyse a possible role of IFN- γ during the anti-tumour response, sporadic tumour development was observed in LoxP-Tag/IFN- γ -deficient and LoxP-Tag/IFN- γ -competent mice. The absence of IFN- γ slightly reduced the tumour latency in about half of the mice, but all mice from both groups developed a tumour (Supplementary Fig. 8) that did not differ in spectrum or number. IFN- γ -deficient Tag-negative mice remained tumour-free during 20 months of observation. Even if the shortened latency in some IFN- γ -deficient compared with IFN- γ -competent mice is due to a tumour-specific response, it is unsuccessful in selecting variants of low immunogenicity (see above). Alternatively, altered pathogen status and inflammatory responses in LoxP-Tag/IFN- γ -deficient mice may explain the reduced tumour latency¹⁸. To explain the IgG2b anti-Tag antibodies, we determined serum TGF- β 1 levels. In tumour-bearing mice, substantially elevated TGF- β 1 concentrations were detected in comparison with young LoxP-Tag and old B6 mice (Fig. 4c). The tumour cells in most (five of six) cases did not produce

TGF- β 1 (Supplementary Fig. 9). This indicates that TGF- β 1 expression might not be a direct escape mechanism by the tumour but might instead be produced by host cells. In tumour-bearing LoxP-Tag mice a generalized and complete IFN- γ deficiency was detected by a sensitive *in vivo* IFN- γ capture assay. Whereas young LoxP-Tag and old B6 mice contained substantial amounts of IFN- γ in serum (500 pg ml⁻¹ or more), virtually no IFN- γ could be detected in the serum of tumour-bearing mice (Fig. 4c). It therefore seems that TGF- β 1 dominates over IFN- γ . It is currently not known when, relative to each other, both cytokines are involved and whether transiently tumoricidal effector T cells are generated.

It is still unclear whether T cells spontaneously reject tumours or select variants of low immunogenicity^{1–10,19–21}. The model of sporadic tumours with a defined tumour-specific rejection antigen offered a unique possibility of analysing the spontaneous immune response against tumours that develop from single cells over a period of several months, thus reflecting physiological tumour development as closely as possible (with the likely exception of their high immunogenicity). We showed that a primary immunogenic tumour is unable to induce or sustain a protective immune response, but instead induces tolerance. In a comparable manner to a study in melanoma patients²², tolerance was accompanied by expansion of anergic CD8⁺T cells. At some point the sporadic tumours seem to transiently induce an IFN- γ -associated response that is generally required for the rejection of transplanted tumours²³. However, this response is inefficient in selecting less immunogenic variants and can only slightly, if at all¹⁸, delay tumour development. We cannot exclude the possibility that a tumour was occasionally rejected, but this is unlikely because IFN- γ -deficient mice did not develop multiple tumours and LoxP-Tag mice that rejected a tumour once—the transplanted tumour—remained tumour-free throughout their life, whereas non-treated mice all developed a tumour. Even though many factors, such as inappropriate dendritic cell activation²⁴, might be responsible for the failure to control the autochthonous tumour, our data can best be explained by 'sneaking through'; for example, the tumour is recognized when it is already refractory to destruction²⁵. The growth of regressor tumours (those that are rejected by syngeneic recipients without previous immunization) has been attributed to immune deficiency of the tumour-bearing host, for example that caused by chronic ultraviolet exposure or T-cell deficiency in Rag-2^{-/-} mice^{6,26}. The lower immunogenicity of tumours from immune-competent mice has therefore previously been explained by the T-cell-mediated selection of a less immunogenic variant⁶. In contrast, in our model, tumours induce T-cell tolerance without losing their intrinsic immunogenicity. Further analysis will be needed to show whether immune selection occurs, when there is no selective pressure for expression of the tumour antigen, and how our data relate to the human system.

METHODS

Mice. For the generation of transgenic mice, plasmid pCAG-CAT-Z²⁷ was digested with *Bam*HI to remove the *lacZ* gene. The *Eco*RV-digested SV40 Tag DNA fragment, derived from plasmid pCMV-Tagori²⁸, was ligated into the filled-in pCAG-CAT plasmid to generate pCAG-CAT-Tag. Transgenic mice (LoxP-Tag) on a C57BL/6 background were obtained by standard methods. C57BL/6 mice (Charles River, Sulzfeld, Germany), Rag-1^{-/-} (B6.129S7-Rag1tm1Mom), Albumin-Cre transgenic (B6.Cg-Tg(Alb-cre)21Mgn/J) and IFN- γ -deficient mice (B6.129S7-Ifngtm1Ts) (The Jackson Laboratory) and SCID mice (Taconics Farms) were used.

Tumour experiments. LoxP-Tag mice showing obvious tumours or other signs of distress were subjected to full necropsy, and with few exceptions a tumour was identified macroscopically. Tumours from LoxP-Tag mice had been cultured for not more than eight passages before 10⁶ cells of the indicated cells were injected subcutaneously into the indicated mice. Animals without tumours were monitored for at least 90 days and scored as tumour-bearing when the tumour size was at least 10 mm in diameter. For liver tumour induction by somatic CAT gene deletion, mice were injected intravenously with 10⁹ plaque-forming units of Cre-expressing adenoviruses. The Ad.Cre vector with the CMV-driven *nlsCre*

gene was constructed by standard methods. IL-7/B7.1 MCA-205 fibrosarcoma²⁹ and B16-F10 cells were infected with an SV40 Tag-expressing or mock retrovirus²⁸. Mice were immunized twice by subcutaneous injection of 2.5 × 10⁵ Tag-transfected irradiated (10 Gy) MCA-IL7/B7.1 and challenged with either Tag⁺ or Tag⁻ B16.F10 cells (2.5 × 10⁵).

Polymerase chain reaction. RT-PCR was performed with standard procedures. The PCR primers used were as follows: 5'- β -actin (5'-TGGATCCTGTGGCATCCATGAACTACAT-3') and 3'- β -actin (5'-AAACGCAGCTCAGTAACAGTCGCGCTAGAA-3'); CAT2 (5'-CAGTCAGTTGCTCAATGTACC-3') and CAT3 (5'-ACTGGTGAACTCACCCA-3'); Tag2 (5'-CCTCTCTGTTTAAAC TTTATCC-3') and Tag732 (5'-ATTGAGGAAAGTTGCCAGG-3'); Tag1146 (5'-ATCCATTCTT CTATGTCAGC-3') and 5'-Tag195 (5'-GGTCTTGAAAGG AGTGCCCTGGGGGA-3') and 3'-Tag195 (5'-CCTCAGTTGCATCCCAGAA GCCTCC-3'). For PCR of genomic DNA, primers 5'-loxP1 (5'-CATCATTTT GGCAAAGAATTCAAGC-3') and 3'-loxP2 (5'-GCAAATTTAAAGCGCTGAT GATCC-3') were used.

Immunostaining. Frozen 6- μ m sections of tumour tissue were mounted on slides, fixed in acetone and stained with haematoxylin/eosin. Antibodies used were fluorescein isothiocyanate (FITC)-labelled mouse anti-SV40 large/small T antigen (Pab 108; BD Pharmingen), pan-cytokeratin (Lu-5; BMA), FITC-labelled anti-mouse IgG1 (A85-1; BD Pharmingen), alkaline phosphatase-conjugated goat anti-rat IgG (Jackson Immunoresearch Laboratories) and isotype-matched control mAbs.

Analysis of anti-Tag antibodies. Blood samples from indicated mice were collected. Enzyme-linked immunosorbent assay (ELISA) plates were coated overnight with 4 μ g ml⁻¹ SV40 Tag protein (Chimerx) and blocked with 2% BSA; serial dilutions of sera were adsorbed on coated plates for 30 min. Bound antibody was detected with biotin-conjugated anti-mouse IgG antibodies (Dianova), avidin-peroxidase conjugate (Sigma), penicillin/streptomycin, minimal essential medium and 2-mercaptoethanol (50 μ M). After five days of culture, cells were harvested, washed twice, and incubated with rat anti-mouse CD16/CD32 (BD Pharmingen) for 30 min on ice. After a single wash, the cells were incubated with phycoerythrin-labelled K^b/IV-tetramers (MHC Tetramer Core Facility at Emory University Vaccine Center, or ProImmune Ltd) and FITC-labelled rat anti-mouse CD8a (53-6.7; BD Pharmingen) for 1 h on ice and analysed with a FACScan (Becton Dickinson) flow cytometer. For the *in vivo* cytotoxicity assay, spleen cells from C57BL/6 mice at a concentration of 10⁷ cells ml⁻¹ in ice-cold PBS were loaded with the H2-K^b-restricted Tag-IV peptide at a final concentration of 1 μ M or left without peptide for 15 min at 37 °C.

Subsequently, the cells were labelled with CFSE in a final concentration of 0.75 μ M (CFSE^{high} population) or 0.075 μ M (CFSE^{low} population) for 15 min at 20 °C, washed once in ice-cold RPMI 1640 medium with 10% FCS and twice in ice-cold PBS. A total of 2 × 10⁷ cells at a 1:1 ratio were injected intravenously into the indicated mice. After 18 h, CFSE-labelled cells in the spleen of recipient mice were analysed by flow cytometry. Killing in naive controls was set to 0%; the specific cytolytic activity was calculated as 100 × (percentage of CFSE^{high} cells/percentage of CFSE^{low} cells).

Determination of serum IFN- γ and TGF- β 1 levels. Serum IFN- γ was assayed with the *in vivo* capture assay (BD Pharmingen) in accordance with the manufacturer's instructions. Serum samples were obtained 18 h after capture antibody injection (intravenous) from individual mice. TGF- β 1 serum levels were determined by ELISA (R&D Systems) in accordance with the manufacturer's instructions.

Received 7 February; accepted 23 June 2005.

1. Burnet, M. Immunological factors in the process of carcinogenesis. *Br. Med. Bull.* **20**, 154–158 (1964).
2. Klein, G. & Klein, E. Immune surveillance against virus-induced tumors and nonrejectability of spontaneous tumors: contrasting consequences of host versus tumour evolution. *Proc. Natl Acad. Sci. USA* **74**, 2121–2125 (1977).
3. Stutman, O. Tumor development after 3-methylcholanthrene in immunologically deficient athymic-nude mice. *Science* **183**, 534–536 (1974).
4. Pehn, R. T. Stimulatory effects of immune reactions upon the growths of untransplanted tumors. *Cancer Res.* **54**, 908–914 (1994).
5. Klein, G., Sjogren, H. O., Klein, E. & Hellstrom, K. E. Demonstration of resistance against methylcholanthrene-induced sarcomas in the primary autochthonous host. *Cancer Res.* **20**, 1561–1572 (1960).
6. Shankaran, V. et al. IFN γ and lymphocytes prevent primary tumour development and shape tumour immunogenicity. *Nature* **410**, 1107–1111 (2001).

7. Street, S. E., Trapani, J. A., MacGregor, D. & Smyth, M. J. Suppression of lymphoma and epithelial malignancies effected by interferon gamma. *J. Exp. Med.* **196**, 129–134 (2002).
8. Dunn, G. P., Bruce, A. T., Ikeda, H., Old, L. J. & Schreiber, R. D. Cancer immunoediting: from immunosurveillance to tumour escape. *Nature Immunol.* **3**, 991–998 (2002).
9. Blankenstein, T. & Qin, Z. Chemical carcinogens as foreign bodies and some pitfalls regarding cancer immune surveillance. *Adv. Cancer Res.* **90**, 179–207 (2003).
10. Qin, Z. & Blankenstein, T. A cancer immunosurveillance controversy. *Nature Immunol.* **5**, 3–4 (2004).
11. Johnson, L. *et al.* Somatic activation of the *K-ras* oncogene causes early onset lung cancer in mice. *Nature* **410**, 1111–1116 (2001).
12. Meuwissen, R., Jonkers, J. & Berns, A. Mouse models for sporadic cancer. *Exp. Cell Res.* **264**, 100–110 (2001).
13. Adams, T. E., Alpert, S. & Hanahan, D. Non-tolerance and autoantibodies to a transgenic self antigen expressed in pancreatic beta cells. *Nature* **325**, 223–228 (1987).
14. Schell, T. D., Knowles, B. B. & Tevethia, S. S. Sequential loss of cytotoxic T lymphocyte responses to simian virus 40 large T antigen epitopes in T antigen transgenic mice developing osteosarcomas. *Cancer Res.* **60**, 3002–3012 (2000).
15. Mullen, C. A., Urban, J. L., Van Waes, C., Rowley, D. A. & Schreiber, H. Multiple cancers. Tumor burden permits the outgrowth of other cancers. *J. Exp. Med.* **162**, 1665–1682 (1985).
16. Beck, C., Schreiber, H. & Rowley, D. Role of TGF-beta in immune-evasion of cancer. *Microsc. Res. Tech.* **52**, 387–395 (2001).
17. Gorelik, L. & Flavell, R. A. Transforming growth factor-beta in T-cell biology. *Nature Rev. Immunol.* **2**, 46–53 (2002).
18. Enzler, T. *et al.* Deficiencies of GM-CSF and interferon gamma link inflammation and cancer. *J. Exp. Med.* **197**, 1213–1219 (2003).
19. Khong, H. T. & Restifo, N. P. Natural selection of tumour variants in the generation of 'tumour escape' phenotypes. *Nature Immunol.* **3**, 999–1005 (2002).
20. Schreiber, H. in *Fundamental Immunology* 5th edn (ed. Paul, W.) 1557–1592 (Raven, Philadelphia, 2003).
21. Staveley-O'Carroll, K. *et al.* Induction of antigen-specific T cell anergy: An early event in the course of tumour progression. *Proc. Natl Acad. Sci. USA* **95**, 1178–1183 (1998).
22. Lee, P. P. *et al.* Characterization of circulating T cells specific for tumour-associated antigens in melanoma patients. *Nature Med.* **5**, 677–685 (1999).
23. Blankenstein, T. The role of tumour stroma in the interaction between tumour and immune system. *Curr. Opin. Immunol.* **17**, 180–186 (2005).
24. van Mierlo, G. J. *et al.* Activation of dendritic cells that cross-present tumour-derived antigen licenses CD8⁺ CTL to cause tumour eradication. *J. Immunol.* **173**, 6753–6759 (2004).
25. Old, L. J., Boyse, E. A., Clarke, D. A. & Carswell, E. A. Antigenic properties of chemically-induced tumors. *Ann. NY Acad. Sci.* **101**, 80–106 (1962).
26. Kripke, M. L. & Fisher, M. S. Immunologic parameters of ultraviolet carcinogenesis. *J. Natl. Cancer Inst.* **57**, 211–215 (1976).
27. Sakai, K., Mitani, K. & Miyazaki, J. Efficient regulation of gene expression by adenovirus vector-mediated delivery of the CRE recombinase. *Biochem. Biophys. Res. Commun.* **217**, 393–401 (1995).
28. Li, L. P., Schlag, P. M. & Blankenstein, T. Transient expression of SV 40 large T antigen by Cre/LoxP-mediated site-specific deletion in primary human tumour cells. *Hum. Gene Ther.* **8**, 1695–1700 (1997).
29. Cayeux, S., Richter, G., Noffz, G., Dörken, B. & Blankenstein, T. Influence of gene-modified (IL-7, IL-4, and B7) tumour cell vaccines on tumour antigen presentation. *J. Immunol.* **158**, 2834–2841 (1997).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank S. Horn, D. Barthel, C. Westen, A. Gaertner and M. Roesch for technical assistance; M. Czéh and T. Kammertoens for discussion; and N. Karamatskos for statistical analysis. This work was supported by grants from the Deutsche Forschungsgemeinschaft and the Deutsche Krebshilfe, Dr Mildred Scheel-Stiftung.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to T.B. (tblank@mdc-berlin.de).

Role of nucleophosmin in embryonic development and tumorigenesis

Silvia Grisendi¹, Rosa Bernardi¹, Marco Rossi², Ke Cheng¹, Luipa Khandker¹, Katia Manova³ & Pier Paolo Pandolfi¹

Nucleophosmin (also known as NPM, B23, NO38) is a nucleolar protein directly implicated in cancer pathogenesis, as the *NPM1* gene is found mutated and rearranged in a number of haematological disorders^{1–5}. Furthermore, the region of chromosome 5 to which *NPM1* maps is deleted in a proportion of *de novo* human myelodysplastic syndromes (MDS)^{6–9}, and loss of chromosome 5 is extremely frequent in therapy-related MDS^{9,10}. NPM is a multifunctional protein^{11–15}, and its role in oncogenesis is controversial as NPM has been attributed with both oncogenic and tumour suppressive functions^{16–19}. To study the function of Npm *in vivo*, we generated a hypomorphic *Npm1* mutant series (*Npm1*^{+/-} < *Npm1*^{hy/hy} < *Npm1*^{-/-}) in mouse. Here we report that Npm is essential for embryonic development and the maintenance of genomic stability. *Npm1*^{-/-} and *Npm1*^{hy/hy} mutants have aberrant organogenesis and die between embryonic day E11.5 and E16.5 owing to severe anaemia resulting from defects in primitive haematopoiesis. We show that *Npm1* inactivation leads to unrestricted centrosome duplication and genomic instability. We demonstrate that Npm is haploinsufficient in the control of genetic stability and that *Npm1* heterozygosity accelerates oncogenesis both *in vitro* and *in vivo*. Notably, *Npm1*^{+/-} mice develop a haematological syndrome with features of human MDS. Our findings uncover an essential developmental role for Npm and implicate its functional loss in tumorigenesis and MDS pathogenesis.

The *NPM1* gene (5q35) is found translocated to different partner genes in a number of human haematopoietic malignancies, such as promyelocytic leukaemia¹, anaplastic large cell lymphoma², acute myeloid leukaemia, chronic myelogenous leukaemia and MDS³. *NPM1* has also been found mutated in approximately 35% of acute myeloid leukaemia cases^{4,5}. Furthermore, deletion of variable portions of the long arm of chromosome 5 (5q) or of the whole chromosome 5, is frequently observed in patients with *de novo* and therapy-related MDS^{6–10} as well as in solid tumours²⁰. These observations strongly implicate NPM in the pathogenesis of human cancer.

NPM modulates diverse molecular functions, among which are ribosome biogenesis^{11,12} and centrosome duplication¹⁵. It modulates, both positively and negatively, the activity of some tumour suppressor genes such as p53 (refs 18, 19) and Arf^{17,21}, and it is upregulated in response to stress stimuli including DNA damage¹³ and hypoxia²². However, the physiological function of NPM *in vivo* has not been defined to date and its role in oncogenesis remains controversial.

We targeted *Npm1* in mouse by replacing exons 2–7 with an in-frame coding sequence for green fluorescent protein (GFP) (Supplementary Fig. 1a). Two independent and correctly targeted embryonic stem cell clones were used to generate *Npm1*^{+/-} mice (Supplementary Fig. 1b, c). Targeting of the gene resulted in complete abrogation of *Npm1* gene expression (Supplementary

Fig. 1d) and concomitant expression of the GFP protein (Supplementary Fig. 1e, f). *Npm1* knockout resulted in embryonic lethality between E11.5 and E12.5 (Supplementary Table 1). Heterozygous mice were viable and appeared to thrive, but the ratio between wild-type and *Npm1*^{+/-} mice was only about 1:1.5 (40.5% and 59.5%, respectively).

Although the morphological analysis of *Npm1*^{-/-} embryos showed relatively well-developed heart, neural tube, otic vesicles, brachial arches and limb buds, the mutant embryos were consistently smaller in size compared with wild-type and heterozygote littermates (Fig. 1a; Supplementary Fig. 1e). Placental structures were smaller but normally developed (Supplementary Fig. 1g). However, we observed a decreased number of embryonic erythroblasts at the interface between maternal and fetal blood (Supplementary Fig. 1h).

The *Npm1*^{-/-} mutants also showed distinct developmental abnormalities. Most noticeably, they displayed deficient anterior brain organogenesis, with complete absence of the eye (Fig. 1a–c). Analysis of E9.5 mutant embryos by whole-mount *in situ* hybridization using neuroectoderm markers revealed that *Bfl* (also known as *Foxg1*) messenger RNA, normally expressed in most of the telencephalic region, was completely absent in *Npm1*^{-/-} embryos (Fig. 1b, upper panels). In contrast, expression of *En2*, which normally encompasses the mesencephalon and anterior metencephalon at E9.5, was preserved in the mutant embryos (Fig. 1b, lower panels).

As early as E9.5, *Npm1*^{-/-} embryos were very pale and the yolk sac appeared devoid of blood supply; by about E12.5, embryos lacked an appreciable fetal liver structure (Fig. 1c). Blood islands were detectable in the *Npm1*^{-/-} yolk sac, although reduced in number (Fig. 1d). Blood vessels in *Npm1*^{-/-} embryos were also properly formed and expressed differentiation markers such as CD34 (Supplementary Fig. 1i). In contrast, the number of haematopoietic precursors in the blood islands of the *Npm1*^{-/-} yolk sacs was greatly reduced (Fig. 1d), and their ability to differentiate into various lineages, particularly towards the erythroid compartment, was profoundly impaired (Fig. 1e).

We next generated a hypomorphic *Npm1* allele, targeting a neomycin (*Neo*) cassette within *Npm1* intron 6 (Supplementary Fig. 2a). Owing to transcriptional interference, homozygosity for the hypomorphic allele reduced *Npm1* expression levels to about 10–30% compared to the wild-type gene (Supplementary Fig. 2b). Homozygous hypomorphic mutants, *Npm1*^{hy/hy}, also died *in utero*, but lethality was considerably delayed (approximately E16.5) compared with the *Npm1*^{-/-} embryos (Supplementary Table 2). *Npm1*^{hy/hy} embryos displayed a similar albeit attenuated phenotype compared to *Npm1*^{-/-} mutants (Supplementary Fig. 2c). Some embryos showed partial development of the telencephalic region (Fig. 1f, upper panels), and most showed some haemoglobinization in both the yolk sac (Fig. 1f, lower panels) and fetal liver (not shown).

¹Cancer Biology and Genetics Program, Department of Pathology, ²Laboratory of Cellular Immunobiology, Department of Medicine, and ³Developmental Biology Program, Molecular Cytology Core Facility, Sloan-Kettering Institute, Memorial Sloan-Kettering Cancer Center, New York, New York 10021, USA.

Taken together, these observations reveal that Npm is required for proper forebrain and blood development.

Npm1^{-/-} embryos showed a noticeable degree of apoptosis, especially in neural tissue, as revealed by staining for activated caspase-3. However, the relative number of proliferating cells was almost comparable to wild-type embryos (Supplementary Fig. 3a, b). Apoptosis was accompanied by a sharp upregulation in p53 protein expression (Fig. 2a), also noticeable in *Npm1*^{hy/hy} embryos (Supplementary Fig. 2d).

In *Npm1*^{-/-} mouse embryonic fibroblasts (MEFs), p53 was not only upregulated but it was also functionally active, as indicated by the concomitant upregulation of p21^{Waf1/Cip1} (Fig. 2b). *Npm1*^{-/-} MEFs underwent premature senescence and growth arrest (Fig. 2c, d). Cell cycle and DNA content analyses showed accumulation of cells with a 4n DNA content and polyploidy (Fig. 2d). This profile is consistent with a tetraploid arrest in G1 phase, as indicated by increased expression of p21^{Waf1/Cip1} and lack of upregulation of G2/M markers, such as phospho-histone H3 and cyclin B (data not shown).

This outcome is characteristic of MEFs undergoing p53-dependent post-mitotic tetraploid cell-cycle arrest²³, which can occur in response to mitotic spindle damage ('mitotic slippage')^{23–25}. Indeed, in *Npm1*^{-/-} cells we detected high rates of aberrant mitotic figures with multiple centrosomes and multi-polar spindles (Fig. 2e). Centrosome amplification (≥ 3 centrosomes per cell) was found in up to 30% of interphase *Npm1*^{-/-} MEFs (Fig. 2f). Supernumerary centrosomes were also observed in sections from *Npm1*^{-/-} embryos (Fig. 2g). Furthermore, among the blood cell precursors isolated and cultured from the yolk sac of *Npm1*^{-/-} embryos ($n = 4$), we detected with high frequency (approximately 30–40%) large multi-nucleated cells with up to four nuclei and numerous centrosomes (Fig. 2h).

As *NPM1* gene dosage is reduced by half in cancer cells harbouring chromosomal rearrangements/deletions at the *NPM1* locus, we next asked whether *Npm1* is haploinsufficient in the control of centrosome duplication and cell proliferation. Early passage *Npm1*^{+/-} and *Npm1*^{hy/hy} MEFs had decreased cell proliferation rates (Fig. 3a), and supernumerary centrosomes were clearly found in heterozygous cells

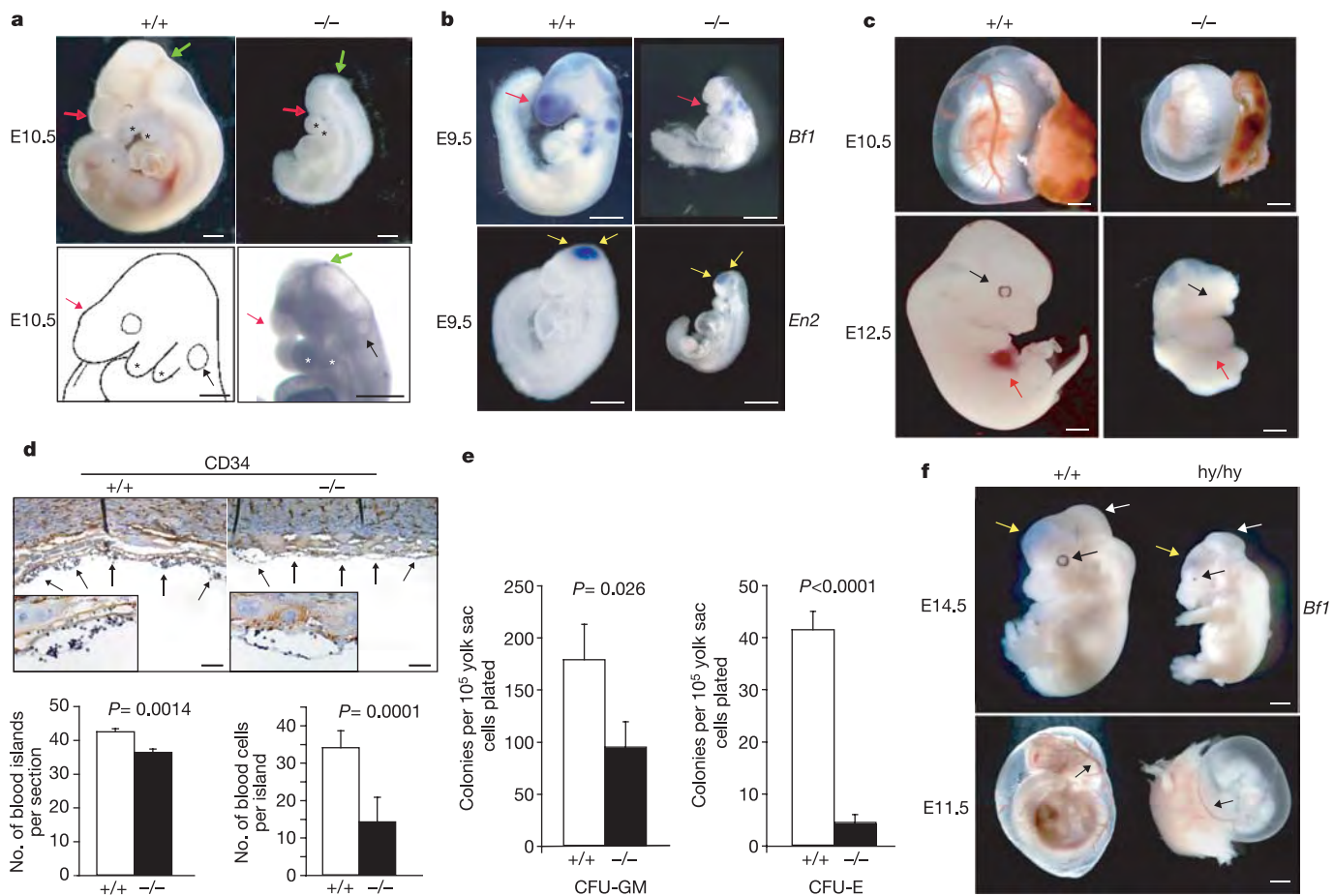


Figure 1 | *Npm1* inactivation causes developmental defects and embryonic lethality. **a**, Whole-mount E10.5 embryos showing that the forebrain is absent (red arrows) and the subdivision between metencephalon and mesencephalon is shifted anteriorly in *Npm1*^{-/-} mutants (green arrows). The lower panels show a comparison between a representative *Npm1*^{-/-} mutant and the graphic representation of a wild-type embryo. Brachial arches (asterisks) and the otic vesicle (black arrows) are well-developed in the mutant. **b**, Whole-mount *in situ* hybridization of E9.5 wild-type (*Npm1*^{+/+}) and *Npm1*^{-/-} embryos with *Bf1* (red arrows) and *En2* (yellow arrows) mRNA probes. **c**, E10.5 wild-type and *Npm1*^{-/-} embryos with intact visceral yolk sacs (upper panels). E12.5 *Npm1*^{-/-} embryos lack the fetal liver (lower panels, red arrows) and the eye (black arrows). **d**, Blood

island formation (arrows) in E9.5 *Npm1*^{+/+} and *Npm1*^{-/-} sections of embryo yolk sac, visualized by CD34 expression (insets show $\times 40$ original magnification). The presence of blood cells is severely compromised in *Npm1*^{-/-} mutants (insets and graphs). **e**, Colony-forming assays from yolk sac progenitors. CFU-GM and CFU-E formation in wild-type and *Npm1*^{-/-} embryo yolk sacs were assayed as described in the Methods ($n = 4$ per group). **f**, *In situ* hybridization of *Bf1* in E14.5 embryos shows partially recovered expression in the *Npm1*^{hy/hy} mutant (yellow arrow), as well as a primordial eye (black arrow). The middle brain is comparable to the wild type in shape and size (white arrow). Lower panels show partial recovery of primitive haematopoiesis in E11.5 *Npm1*^{hy/hy} embryos (black arrow). Scale bars, 500 μ m (**a**, **b**, **c**, **f**), 50 μ m (**d**). Error bars in **d**, **e** represent s.d.

and more frequently observed in hypomorphic cells (Fig. 3b). Furthermore, chromosome fluorescence *in situ* hybridization (FISH) analysis in our hypomorphic series revealed high levels of tetraploidy and aneuploidy in a dose-dependent manner (Fig. 3c; Supplementary Fig. 4a). Thus, *Npm1* gene dosage is a critical determinant in maintaining proper centrosome number and genomic stability.

After long-term culture of MEFs using a 3T9 protocol, we observed that the proliferative rates of *Npm1*^{+/-} MEFs were lower than wild-type cells at early passages (P2–P6) (Fig. 3d, left panel; Fig. 3a). However, a few passages later (P5–P8), *Npm1*^{+/-} cells invariably overcame senescence (data not shown) and showed an immortal phenotype with noticeably higher proliferation rates. In contrast, wild-type MEFs grew faster at early passages but much slower than *Npm1*^{+/-} MEFs upon immortalization (Fig. 3d, left panel). At later passages, *Npm1*^{+/-} MEFs were characterized by higher aneuploidy than wild-type cells (Fig. 3d, right panels), and showed a greater ability to form transformation foci in low-density seeding assays (Fig. 3e). Therefore, the underlying genetic instability shown by *Npm1*-deficient cells may be a driving force for faster acquisition of additional genetic mutations required initially for immortalization (for example, p53 loss) and then transformation. In

a p53 (*Trp53*) null background, not only were *Npm1*^{+/-} MEFs fully rescued in their growth defect at early passage (P3–P5) but they grew faster than *Npm1*^{+/+} MEFs from littermate controls (Supplementary Fig. 4b), and consistently showed higher efficiency at forming foci in low-density seeding assays (data not shown). This further corroborates the notion that the growth disadvantage of *Npm1*-deficient cells is due to upregulation of a functional p53 protein. However, p53 inactivation in *Npm1*^{-/-}p53^{-/-} double mutants did not rescue embryonic lethality, which was nevertheless substantially delayed (Supplementary Table 3). This indicates that the developmental failure observed in *Npm1* mutants is in part, but not solely, due to an acute p53 response.

In soft agar assays, low passage *Npm1*^{+/-} MEFs (P4) were not transformed to the same extent as *Npm1*^{+/+} controls by the combination of Ras^{V12} and E1A oncogenes when cells were assayed immediately after infection/selection (Supplementary Fig. 4c). However, if the same infected cells were further cultured (5–10 passages) before soft agar plating, *Npm1*^{+/-} cells reproducibly yielded approximately 50% more foci than control cells (Supplementary Fig. 4c). A similar outcome was obtained upon infection of already immortalized wild-type and *Npm1*^{+/-} MEFs (P15) with single oncogenes (for example, E1A or c-Myc; Fig. 3f).

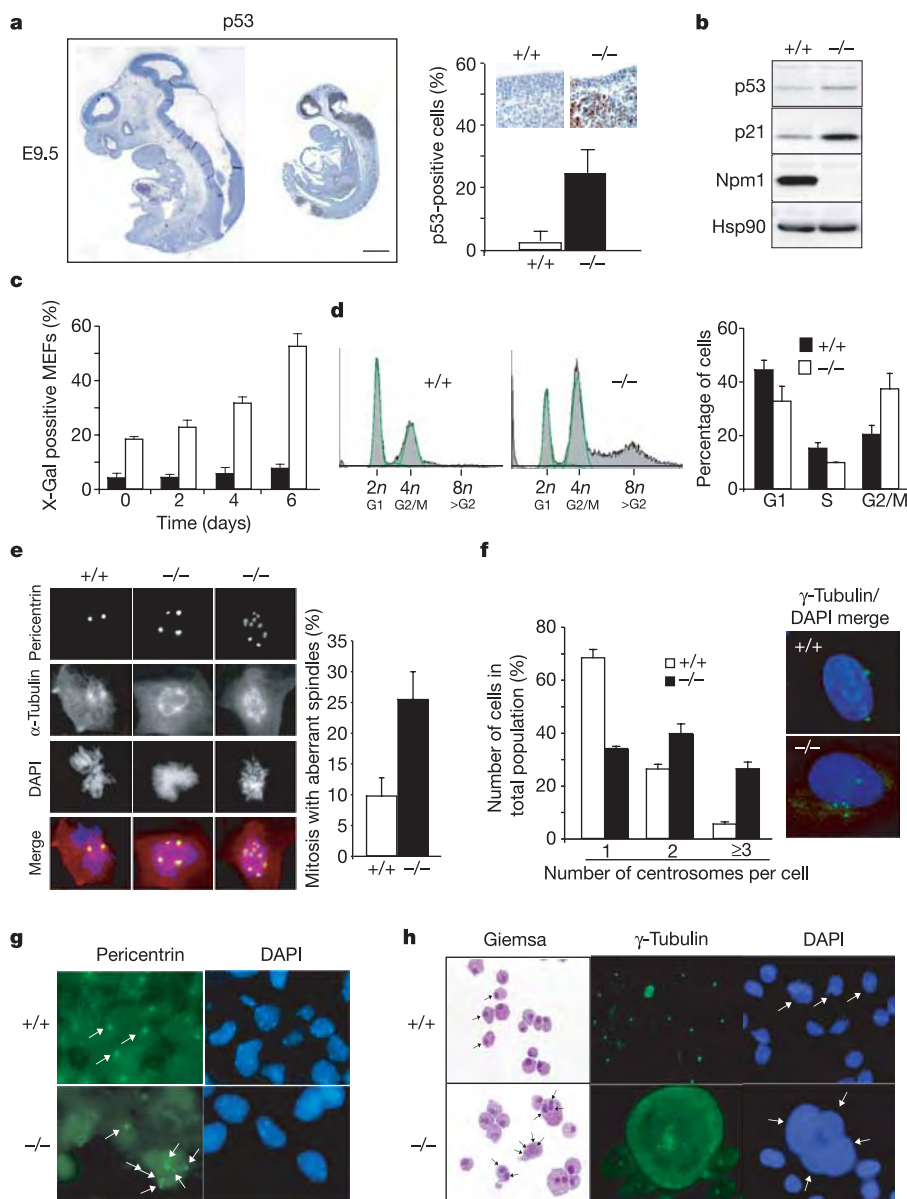


Figure 2 | Effect of *Npm1* inactivation on apoptosis, cell cycle and centrosome number.

a, Histological analysis of p53 expression in *Npm1*^{+/+} and *Npm1*^{-/-} embryos. Scale bar, 500 μm. The graph shows the percentage of p53-expressing cells in the total population of neural tissue (mean ± s.d. of two sections from one embryo of each genotype). **b**, Western blot analysis of p53 and p21^{Waf1/Cip1} expression in *Npm1*^{+/+} and *Npm1*^{-/-} MEFs. **c**, Senescence analysis in early passage (P2) *Npm1*^{+/+} (black) and *Npm1*^{-/-} (white) MEFs. **d**, Cell-cycle profile of P2 *Npm1*^{+/+} and *Npm1*^{-/-} MEFs, showing their DNA content distribution (2n, 4n and 8n). The graph on the right shows percentages of cells in each phase of the cell cycle. **e**, Aberrant mitosis and spindle abnormalities in *Npm1*^{-/-} MEFs, identified by co-immunofluorescence with anti-pericentrin and α-tubulin antibodies (for centrosomes and mitotic spindles, respectively). Graph shows percentage of abnormal mitosis over total mitosis for *Npm1*^{+/+} and *Npm1*^{-/-} mice. **f**, Percentages of cells with the indicated number of centrosomes in P2 *Npm1*^{+/+} and *Npm1*^{-/-} MEFs. Values represent the average of three independent experiments. Representative *Npm1*^{+/+} and *Npm1*^{-/-} cells stained with an anti-γ-tubulin antibody (green) and DAPI (blue) are shown. **g**, Centrosome amplification in E10.5 *Npm1*^{-/-} embryo tissue sections revealed by pericentrin immunostaining. **h**, Giemsa and DAPI staining reveal multinucleated *Npm1*^{-/-} cells (arrows), and γ-tubulin staining shows centrosome amplification in yolk sac progenitor cells from E9.5 *Npm1*^{-/-} embryos. In all experiments, pericentrin and γ-tubulin antibodies were used interchangeably since they consistently gave similar results. Results are presented as mean ± s.d. from three or four independent experiments.

As *c-Myc* transformation potential was markedly enhanced by *Npm1* heterozygosity (Fig. 3f), we investigated B-cell lymphoma onset in transgenic mice expressing *c-Myc* under the control of the immunoglobulin heavy chain enhancer (*Eμ-Myc*). Disease onset in *Npm1*^{+/-}*Eμ-Myc* mutant mice was significantly accelerated compared with *Npm1*^{+/+}*Eμ-Myc* mice (Fig. 3g), indicating that loss of one *Npm1* allele cooperates with oncogenes in transformation and tumorigenesis both *in vitro* and *in vivo*.

Genomic DNA analysis of the lymphomas by comparative genomic hybridization (CGH array) revealed profound differences between *Npm1*^{+/+}*Eμ-Myc* and *Npm1*^{+/-}*Eμ-Myc* animals. All of the lymphomas (4/4) analysed from *Npm1*^{+/-}*Eμ-Myc* mice had numerical chromosomal abnormalities, but only one out of the four tumours analysed from *Npm1*^{+/+}*Eμ-Myc* mice showed numerical abnormalities (Supplementary Fig. 4d). In contrast, all of the tumours from *Npm1*^{+/+}*Eμ-Myc* mutants harboured recurrent structural abnormalities, but only one out of the four tumours analysed from *Npm1*^{+/-}*Eμ-Myc* mice showed structural abnormalities

(Supplementary Fig. 4d). Thus, heterozygous inactivation of *Npm1* influences the type of genetic lesion that accumulate in the tumours. Furthermore, CGH array, Southern blot and immunohistochemical analyses did not reveal loss of heterozygosity (LOH) for *Npm1* (Supplementary Fig. 4d–f).

As *Npm1*^{-/-} embryos show dramatic defects in primitive haematopoiesis and *Npm1* proved to be haploinsufficient, we analysed the haematopoiesis of *Npm1*^{+/-} mutants. *Npm1*^{+/-} mice showed features found in human MDS. Approximately 80% of the mice analysed at 6–18 months of age (*n* = 22) showed one or more of these features (Table 1). Peripheral blood from *Npm1*^{+/-} mice showed significantly increased mean corpuscular volume (MCV; *P* = 0.02) and red cell distribution width (RDW; *P* = 0.0089) values (Fig. 4a and Table 1), indicating an overall increase in size and anisocytosis of the erythrocytes. This was not accompanied by reticulocytosis (Fig. 4a), (*Npm1*^{+/+} = 2.95 ± 0.8%, *n* = 6; *Npm1*^{+/-} = 3.66 ± 1.6%, *n* = 11), with very low reticulocyte counts found in a few *Npm1*^{+/-} mutants (Fig. 4a). Red blood

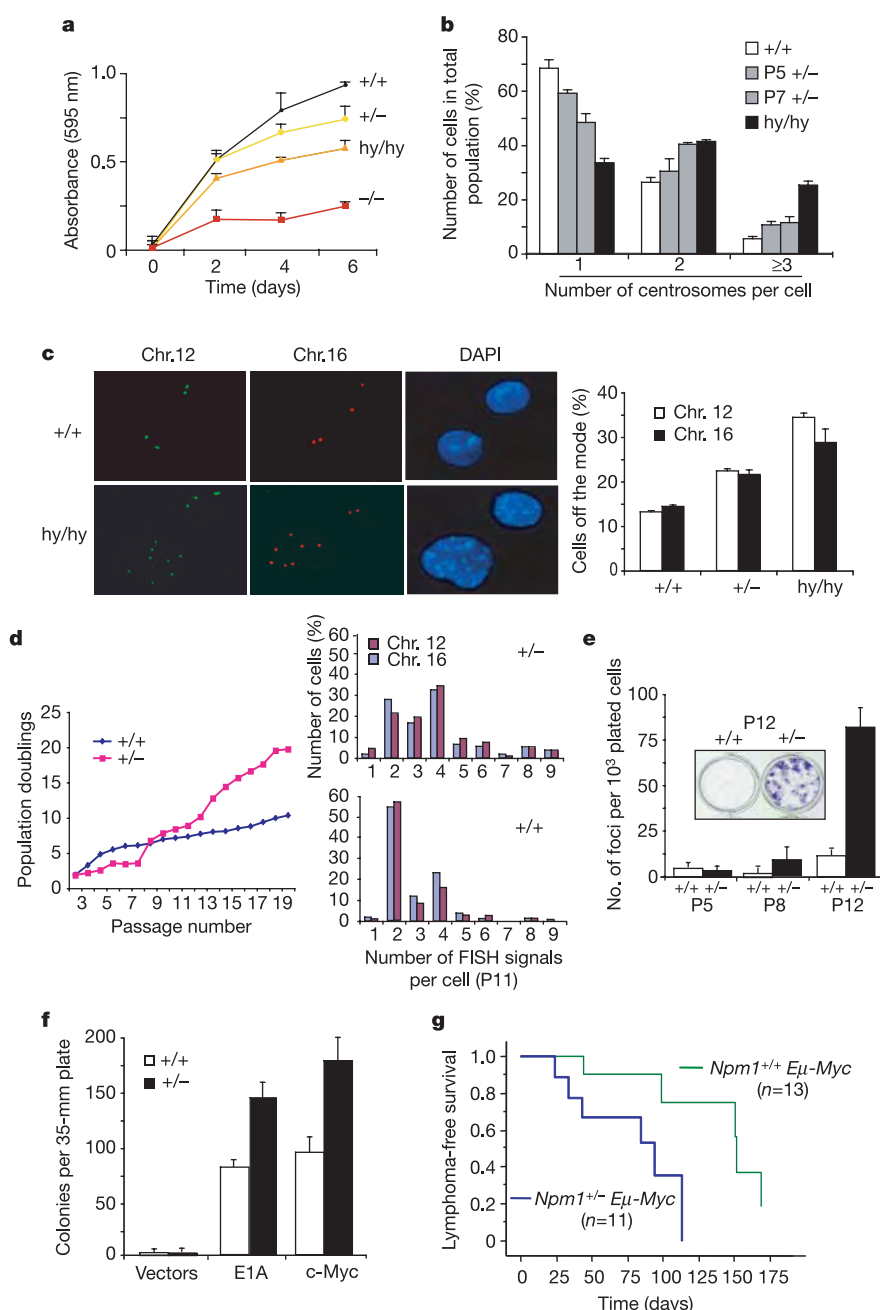


Figure 3 | *Npm1* is haploinsufficient for the control of centrosome duplication and maintenance of genomic stability.

a, Growth curves of the *Npm1*^{+/+} < *Npm1*^{+/-} < *Npm1*^{hy/hy} < *Npm1*^{-/-} hypomorphic MEFs series (see Methods). **b**, Percentages of cells with normal (1 and 2) and aberrant (≥3) number of centrosomes in MEFs, with genotype and passage number indicated. **c**, Genomic instability in *Npm1*-deficient cells: chromosome FISH analysis with probes specific for mouse chromosomes 12 (green) and 16 (red) of *Npm1*^{+/+} and *Npm1*^{hy/hy} MEFs. Percentages of *Npm1*^{+/+}, *Npm1*^{+/-} and *Npm1*^{hy/hy} cells with a number of FISH signals different from the modal value of 2 are shown. **d**, Proliferation rates of *Npm1*^{+/+} and *Npm1*^{+/-} MEFs during serial passaging, shown as population doublings. Results show the average of 8 different MEF preparations per genotype, passaged on a 3T9 protocol. Right panels show a representative chromosome FISH analysis of P11 *Npm1*^{+/+} and *Npm1*^{+/-} MEFs. **e**, Low-density seeding focus-forming assay on *Npm1*^{+/+} and *Npm1*^{+/-} MEFs at different passages (described in Methods). Inset shows a representative image of the results obtained at P12. **f**, Soft agar colony-forming assay on P15 *Npm1*^{+/+} and *Npm1*^{+/-} MEFs upon infection with E1A or *c-Myc* oncogenes. Empty vectors were used as control. **g**, B-cell lymphoma-free survival in mice of the indicated genotype (*P* = 0.0202 by Mantel-Cox log-rank analysis of the two genotypes). Results are presented as mean ± s.d.

Table 1 | Summary of haematopoietic characteristics of *Npm1*^{+/+} and *Npm1*^{+/-} mice

Mice	No. of mice analysed*	Elevated MCV†	Elevated RDW‡	Abnormal platelet counts§	Erythroid morphological dysplasia	Dysplastic megakaryocytes¶	Defective erythroid maturation with hypercellularity#
<i>Npm1</i> ^{+/+}	19	2/7	1/7	2/7	0/16	3/12	1/8
<i>Npm1</i> ^{+/-}	22	6/8	5/9	8/9	12/16	7/11	6/9

*Mice aged 6–18 months.
†MCV was scored as elevated when higher than 56.0 femtolitres (average in *Npm1*^{+/+} mice is 55.0 femtolitres).
‡RDW was scored as elevated when higher than 14% (average in *Npm1*^{+/+} mice is 12.95%).
§Platelet counts were considered abnormal when higher than $1,200 \times 10^3$ and lower than $1,000 \times 10^3$ per μl of blood (average in *Npm1*^{+/+} mice is $1,105 \times 10^3 \mu\text{l}^{-1}$).
|| A mouse was scored positive when the fraction of dysplastic cells was higher than 5% of the total number of erythroid precursors (an average of 0–5% dysplastic cells was observed in wild-type mice).
¶ A mouse was scored positive when the fraction of hypobulbated megakaryocytes and megakaryocytes with multiple small nuclei was higher than 30% (0–30% dysplastic cells were observed in wild-type mice).
Average values are shown in Supplementary Fig. 5c.

cell (RBC) counts and haemoglobin (HGB) levels of *Npm1*^{+/-} mice ($n = 14$) did not differ from the wild-type littermates ($n = 11$) (RBC: *Npm1*^{+/+} = $9.52 \pm 0.5 \times 10^6 \mu\text{l}^{-1}$, *Npm1*^{+/-} = $9.40 \pm 1.1 \times 10^6 \mu\text{l}^{-1}$; HGB: *Npm1*^{+/+} = $14.75 \pm 1.7 \text{ g dl}^{-1}$, *Npm1*^{+/-} = $14.77 \pm 2.7 \text{ g dl}^{-1}$). In a few *Npm1*^{+/-} mice thrombo-

cyte numbers were markedly increased, and in general the platelet counts showed a wider scatter in the mutants compared with wild-type littermates (Fig. 4a and Table 1; *Npm1*^{+/+} = $1,105 \pm 99 \times 10^3 \mu\text{l}^{-1}$, *Npm1*^{+/-} = $1,282 \pm 630 \times 10^3 \mu\text{l}^{-1}$). No differences were observed in the levels of plasma erythropoietin

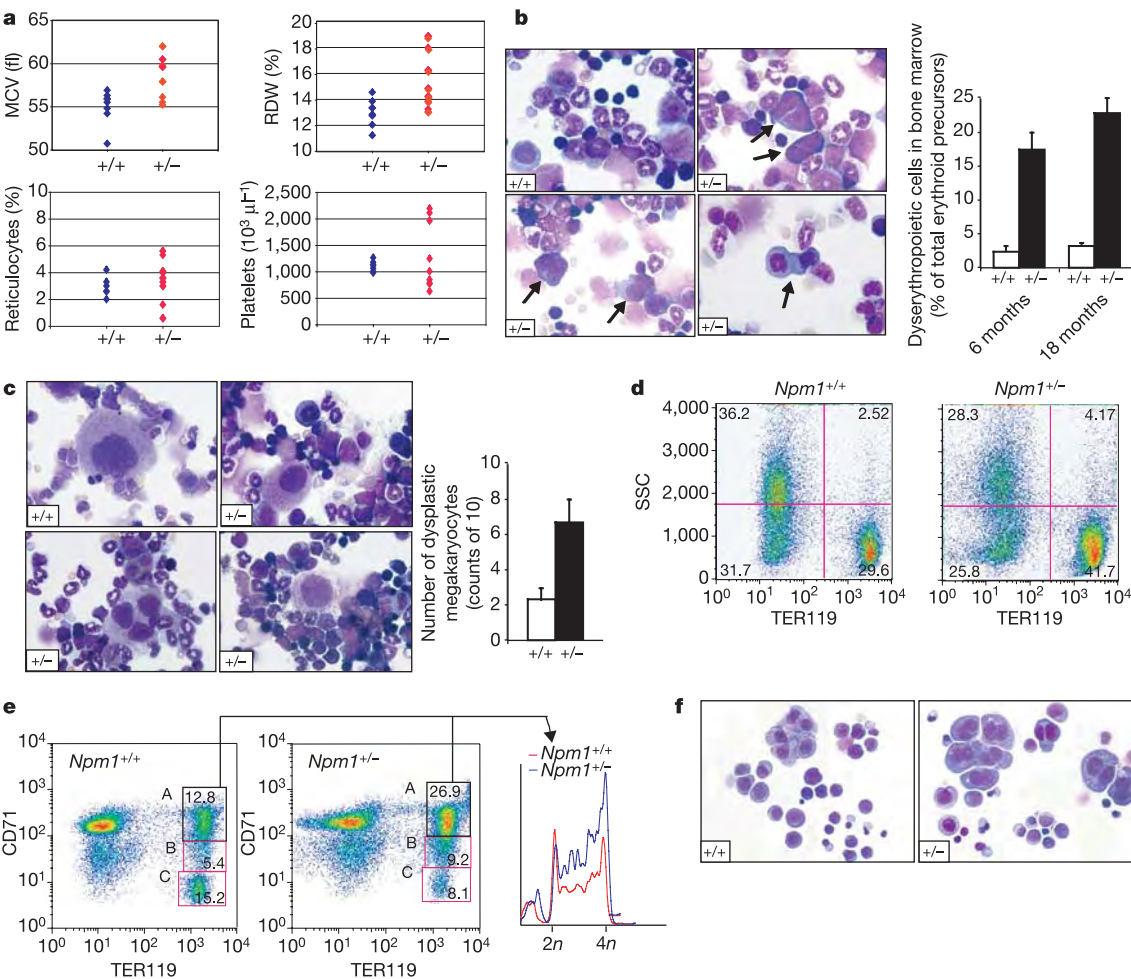


Figure 4 | *Npm1*^{+/-} mice show myelodysplastic features. **a**, Peripheral blood from *Npm1*^{+/-} mice has elevated MCV and RDW values compared to *Npm1*^{+/+} littermates. Reticulocyte and platelet counts are shown in lower panels. **b**, Dyserythropoiesis in *Npm1*^{+/-} mice. Representative images of bone marrow samples from one *Npm1*^{+/+} and one *Npm1*^{+/-} mouse; arrows point to binucleated erythroid cells and abnormal mitotic figures. Histograms show the incidence of dyserythropoietic cells as percentage of the total erythroid cell population (mean $\pm$ s.d.). A total of 16 sex- and age-matched *Npm1*^{+/+} and *Npm1*^{+/-} mice were analysed. For each mouse, 5 independent counts of total erythroid cells ($n = 100$) were performed. **c**, Dysplastic megakaryocytes. Representative bone marrow samples from *Npm1*^{+/+} and *Npm1*^{+/-} mice show megakaryocytes with a small oval

nucleus in mature cytoplasm, multiple separated nuclei or a hypobulbated nucleus in *Npm1*^{+/-} mice. Histogram shows the number of dysplastic megakaryocytes per counts of 10 cells (mean $\pm$ s.d.). A total of 11 *Npm1*^{+/+} mice and 12 *Npm1*^{+/-} mice were analysed; for each mouse, 5 independent counts were performed. **d**, Bone marrow FACS analysis of TER119-expressing cells in *Npm1*^{+/+} and *Npm1*^{+/-} mice. **e**, Increased percentage of TER119^{hi}CD71^{hi}-expressing (boxes A and B) and decreased percentage of TER119^{hi}CD71^{low}-expressing (box C) cells in *Npm1*^{+/-} mouse bone marrow. Cell-cycle analysis of cell population in boxes A (TER119^{hi}CD71^{hi}) is shown on the right. **f**, Representative images of cytopins from CFU-E/BFU-E colonies generated in *in vitro* colony-forming assays using *Npm1*^{+/+} and *Npm1*^{+/-} bone marrow.

($Npm1^{+/+} = 127.12 \pm 13.5 \text{ pg ml}^{-1}$, $n = 6$; $Npm1^{+/-} = 126.5 \pm 20.1 \text{ pg ml}^{-1}$, $n = 11$).

In the bone marrow of $Npm1^{+/-}$ mice, we detected the presence of dyserythropoiesis. We found a high proportion of dysplastic erythroid precursors, among which were binucleated cells, cells with abnormal mitosis (Fig. 4b and Table 1) and cells with nuclear fragmentation and irregular nuclear contours (Supplementary Fig. 5a). Megakaryocyte dyspoiesis was observed in a fraction of these mice (Fig. 4c and Table 1): cells with multiple separated nuclei, with small oval nuclei in mature cytoplasm or with hypolobated nuclei. In a few mutants we also found a marked increase in the number of megakaryocytes (Supplementary Fig. 5b).

The bone marrow of $Npm1^{+/-}$ mice showed hypercellularity of RBC precursors, evaluated as expression of the erythroid surface marker TER119 (ref. 26; Fig. 4d, Table 1 and Supplementary Fig. 5c) and an increased relative ratio of erythroblast/myeloid precursors ($Npm1^{+/+} = 0.39 \pm 0.06$, $n = 6$; $Npm1^{+/-} = 0.55 \pm 0.20$, $n = 6$; Supplementary Fig. 5d). $Npm1^{+/-}$ bone marrow also showed an increased relative percentage of TER119^{hi}CD71^{hi}-expressing cells (markers that identify an immature erythroblast population; boxes A and B in Fig. 4e), and a decrease in TER119^{hi}CD71^{low}-expressing cells (indicating more advanced stages of erythroid differentiation; box C in Fig. 4e, Table 1 and Supplementary Fig. 5c). Cell-cycle analysis of TER119^{hi}CD71^{hi} precursors revealed a significant increase in the fraction of tetraploid cells in the $Npm1^{+/-}$ mutants (Fig. 4e, right panel), as previously observed in $Npm1^{+/-}$, $Npm1^{hy/hy}$ and $Npm1^{-/-}$ cells. Furthermore, although comparable numbers of granulocyte/macrophage colony-forming units (CFU-GM) and erythroid colony/burst-forming units (CFU-E/BFU-E) were obtained in bone marrow *in vitro* colony-forming assays (not shown), the $Npm1^{+/-}$ CFU-E/BFU-E colonies often contained binucleated cells (5–10%; Fig. 4f). Together, these findings demonstrate the presence of dysplastic features in the $Npm1^{+/-}$ mice.

In this work, we attribute to Npm an essential role in embryogenesis. Development of the forebrain and haematopoiesis are profoundly and selectively disrupted in $Npm1$ mutants. These defects may be in part due to cell death resulting from centrosome abnormalities, although defects in ribosome biogenesis observed in $Npm1$ -deficient cells (Supplementary Fig. 2e) might also have a role²⁷.

We demonstrate that Npm function is required to maintain genomic stability through the regulation of centrosome duplication. We propose a model by which Npm functional inactivation causes unrestrained centrosome duplication, in turn leading to aberrant centrosome numbers and aneuploidy. We find that $Npm1$ is haploinsufficient for this function as $Npm1$ heterozygous cells show a significant degree of genomic instability, resulting in increased susceptibility to oncogenic transformation both *in vitro* and *in vivo*. This is relevant not only because of the direct involvement of the *NPM1* locus in human cancer, but also because the region to which *NPM1* maps to is often lost in malignancies^{6–10,20}.

We implicate Npm in the pathogenesis of human MDS. $Npm1^{+/-}$ mice show some of the features found in the human syndrome, such as erythroid and megakaryocytic dysplasia. Deletion of 5q, where *NPM1* is located, or of the whole chromosome 5, is a common structural aberration in patients with *de novo* or therapy-related MDS. On the basis of our findings, functional loss of NPM might contribute to the multifaceted nature of human MDS.

METHODS

Generation of *Npm1* mutant mice. For the *Npm1* knockout construct, 5' (4.2 kb) and 3' (4.1 kb) arms of the targeting vector were obtained by polymerase chain reaction (PCR) amplification from 129/Sv mouse genomic DNA. Exogenous restriction sites were added at the end of the PCR products (5' arm, *PmeI*-*NcoI*; 3' arm, *PacI*-*AscI*), fragments were fully sequenced and subsequently subcloned into a pBlueScript KS vector (Stratagene) endowed with an accordingly modified polylinker. The GFP complementary DNA from pEGFP-N1

(Clontech) was inserted in-frame with exon 2 at the 5' arm end, and a modified ACN cassette²⁸ (pACE-*Cre*-pGK-*Neo*, courtesy of M. Capecchi) was inserted between the GFP sequence and the 3' arm.

For the hypomorphic allele construct, one *EcoRI*-*EcoRI* 8-kb fragment and one *DraI*-*PacI* 4.6-kb fragment of the *Npm1* gene were obtained by PCR amplification from 129/Sv mouse genomic DNA, and subcloned into the pKOII plasmid²⁹ (provided by N. Bardeesy and R. A. Depinho) at *SmaI* and *HpaI* sites respectively, after appropriate modification. The targeting plasmids were linearized and transfected into embryonic stem cells with a 129/Sv background (CJ7). Homologously recombined embryonic stem cell clones were injected into C57BL/6 blastocysts to generate chimaeric progeny. Male chimaeras were backcrossed to C57BL/6 females and germline transmission in F₁ heterozygous offspring was verified by Southern blot analysis. F₁ heterozygotes were interbred to obtain homozygous mutant mice.

Histological and whole-mount *in situ* hybridization analysis of *Npm1* mutant embryos. Embryos at different stages were fixed in 4% paraformaldehyde and dehydrated for whole-mount observation, or for embedding in paraffin to generate 8- μm sections. Sections were stained with haematoxylin and eosin according to standard procedures, or processed for immunostaining. Immunohistochemical detection was performed using an automated staining processor (Discovery, Ventana Medical Systems, Inc.). The staining protocols were established at the Molecular Cytology Core Facility. The primary antibodies were directed against NPM (18-7288, Zymed; 1:500), Ki67 (NCL-Ki67p, Novocastra; 0.16 $\mu\text{g ml}^{-1}$), activated caspase-3 (9661, Cell Signalling Technology; 1 $\mu\text{g ml}^{-1}$), p53 (sc-6243, Santa Cruz Biotechnology; 1 $\mu\text{g ml}^{-1}$), GFP (6556, Abcam; 0.5 $\mu\text{g ml}^{-1}$) and CD34 (09431D, Pharmingen, 2.5 $\mu\text{g ml}^{-1}$). mRNA *in situ* hybridization with digoxigenin-labelled RNA probes was performed in whole-mount embryonic preparations. An InSituPro (Intavis AG) staining processor was used.

Immunofluorescence. Cells grown on coverslips were fixed in cold methanol for 10 min and then briefly incubated in iced acetone. The antibodies used for immunofluorescence microscopy were anti-mouse γ -tubulin (T-6557, Sigma; 1:1,000), anti-rabbit γ -tubulin (T-5192, Sigma; 1:1,000), anti- α -tubulin (T-9026, Sigma; 1:1,000), anti-pericentrin (PRB-432C, Covance; 1:1,000) and anti-NPM (18-7288, Zymed, 1:1,500). For detection, cells were incubated with FITC-conjugated and/or Cy5-conjugated anti-mouse and anti-rabbit secondary antibodies (Jackson ImmunoResearch).

Cell culture, cell-based assays and western blot analysis. MEFs were prepared from E10.5–E13.5 embryos. Early passage MEFs (P2–5) were used in all experiments except as indicated. For cell-cycle analysis, cells were fixed in 70% ethanol, stained with propidium iodide and analysed by flow cytometry. For proliferation analysis, growth curves were generated by seeding 2.5×10^4 cells per well in 12-well plates, each clone in triplicate. Plates were fixed on days 0, 2, 4 and 6 except as indicated. Fixed plates were stained with crystal violet, extracted with 10% acetic acid, and the relative cell number was measured by absorbance at 595 nm. For 3T9 analysis, 9×10^5 cells were passaged into 10-cm dishes every 3 days. A culture was considered 'senescent' when the number of cells decreased (that is, $< 9 \times 10^5$) over two consecutive passages.

Focus-forming assays with $Npm1^{+/+}$ and $Npm1^{+/-}$ MEFs were performed by plating a total of 3×10^3 cells per well in 12-well plates (in triplicate) and culturing for 2 weeks before staining with crystal violet. Soft agar colony-forming assays were performed with $Npm1^{+/+}$ and $Npm1^{+/-}$ MEFs (passage number as indicated) infected with empty control, pBabe-Puro-Ras^{V12}, pBabe-Hygro-E1A and pBabe-Puro-Myc retroviruses, separately or combined as indicated. Transfections of the packaging cell line (Phoenix) were done using Lipofectamine Plus in Optimem (Gibco) without serum or antibiotics. Upon infection, selection and harvesting of the MEFs, transformation assays were performed by resuspending 5×10^4 cells in complete DMEM (2 $\times$) containing 0.3% agarose and seeding (in triplicate) into 35-mm plates containing a 3-ml layer of solidified 0.6% agar in complete medium. Foci were scored 15–20 days later.

For the analysis of senescent cells, staining for the senescence marker β -Gal activity (X-Gal) was performed according to the manufacturer's protocol (Senescence Detection Kit, Oncogene Research Products). For western blot analysis, the following antibodies were used: anti-NPM (18-7288, Zymed), anti-p53 (CM5, Novocastra) and anti-p21 (clone F5, Santa Cruz Biotechnology).

Analysis of haematopoiesis in *Npm1*^{+/-} mice. Peripheral blood collected by tail or retro-orbital bleeding of wild-type and $Npm1^{+/-}$ mice was (1) smeared and stained with May-Grunwald-Giemsa, (2) stained with new methylene blue and smeared for reticulocyte analysis, or (3) subjected to an automated blood analyser (ADVIA 120, Hematology System). Morphological analysis and cell counting of bone marrow samples were performed on cytopins (5×10^5 total

cells) stained with May-Grunwald-Giemsa. For the surface labelling of bone marrow cells and fluorescence-activated cell sorting (FACS) analysis we used APC-conjugated rat anti-mouse TER119 (116211, BioLegend) and PE-conjugated rat anti-mouse CD71 (113807, BioLegend) antibodies. Antibody-labelled single-cell suspensions were analysed by FACScan (Becton Dickinson) and sorted for subsequent propidium iodide staining using MoFlo (DAKO Cytomation).

Chromosome FISH analysis. FISH analysis was performed using chromosome-specific probes. The probes were bacterial artificial chromosomes from a mouse 1-Mb clone set as described³⁰. The chromosome 12 probe (MMU12) corresponds to RP23-54G4, and the chromosome 16 probe (MMU16) corresponds to RP23-290E4. The probes were labelled with biotin or digoxigenin by nick translation. Slides were pre-treated with pepsin before hybridization to ensure maximum hybridization efficiency, and counterstained with DAPI for microscopic analysis and counting.

CGH array analysis. Genomic DNA extracted from lymphomas of four *Npm1*^{+/+}*Eμ-Myc* and four *Npm1*^{+/−}*Eμ-Myc* mice was subjected to comparative genomic hybridization array analysis at the MSKCC Genomics Core Lab (CGL) Microarray Facility. For each mouse, genomic DNA extracted from the tail was used as a reference.

Received 23 February; accepted 14 June 2005.

Published online 6 July 2005.

- Redner, R. L., Rush, E. A., Faas, S., Rudert, W. A. & Corey, S. J. The t(5;17) variant of acute promyelocytic leukemia expresses a nucleophosmin-retinoic acid receptor fusion. *Blood* **87**, 882–886 (1996).
- Morris, S. W. *et al.* Fusion of a kinase gene, *ALK*, to a nucleolar protein gene, *NPM*, in non-Hodgkin's lymphoma. *Science* **263**, 1281–1284 (1994).
- Yoneda-Kato, N. *et al.* The t(3;5)(q25.1;q34) of myelodysplastic syndrome and acute myeloid leukemia produces a novel fusion gene, *NPM-MLF1*. *Oncogene* **12**, 265–275 (1996).
- Falini, B. *et al.* Cytoplasmic nucleophosmin in acute myelogenous leukemia with a normal karyotype. *N. Engl. J. Med.* **352**, 254–266 (2005).
- Grisendi, S. & Pandolfi, P. P. *NPM* mutations in acute myelogenous leukemia. *N. Engl. J. Med.* **352**, 291–292 (2005).
- Van den Berghe, H. & Michaux, L. 5q-, twenty-five years later: a synopsis. *Cancer Genet. Cytogenet.* **94**, 1–7 (1997).
- Westbrook, C. A. *et al.* Cytogenetic and molecular diagnosis of chromosome 5 deletions in myelodysplasia. *Br. J. Haematol.* **110**, 847–855 (2000).
- Giagounidis, A. A. N. *et al.* Clinical, morphological, cytogenetic, and prognostic features of patients with myelodysplastic syndromes and del(5q) including band q31. *Leukemia* **18**, 113–119 (2004).
- List, A. F., Vardiman, J., Issa, J. P. & DeWitte, T. M. Myelodysplastic syndromes. *Hematology (Am. Soc. Hematol. Educ. Program)*, 297–317 (2004).
- Olney, H. J. & Le Beau, M. M. in *The Myelodysplastic Syndromes, Pathobiology and Clinical Management* (ed. Bennet, J. M.) 89–120 (Marcel Dekker, New York, 2002).
- Hingorani, K., Szebeni, A. & Olson, M. O. Mapping the functional domains of nucleolar protein B23. *J. Biol. Chem.* **275**, 24451–24457 (2000).
- Savkur, R. S. & Olson, M. O. Preferential cleavage in pre-ribosomal RNA by protein B23 endoribonuclease. *Nucleic Acids Res.* **26**, 4508–4515 (1998).
- Wu, M. H. & Yung, B. Y. UV stimulation of nucleophosmin/B23 expression is an immediate-early gene response induced by damaged DNA. *J. Biol. Chem.* **277**, 48234–48240 (2002).
- Zhang, H. *et al.* B23/nucleophosmin serine 4 phosphorylation mediates mitotic functions of polo-like kinase 1. *J. Biol. Chem.* **279**, 35726–35734 (2004).
- Okuda, M. *et al.* Nucleophosmin/B23 is a target of CDK2/cyclin E in centrosome duplication. *Cell* **103**, 127–140 (2000).
- Kondo, T. *et al.* Identification and characterization of nucleophosmin/B23/numatrin which binds the anti-oncogenic transcription factor IRF-1 and manifests oncogenic activity. *Oncogene* **15**, 1275–1281 (1997).
- Bertwistle, D., Sugimoto, M. & Sherr, C. J. Physical and functional interactions of the Arf tumour suppressor protein with nucleophosmin/B23. *Mol. Cell. Biol.* **24**, 985–996 (2004).
- Colombo, E., Marine, J. C., Danovi, D., Falini, B. & Pelicci, P. G. Nucleophosmin regulates the stability and transcriptional activity of p53. *Nature Cell Biol.* **4**, 529–533 (2002).
- Kurki, S. *et al.* Nucleolar protein NPM interacts with HDM2 and protects tumour suppressor protein p53 from HDM2-mediated degradation. *Cancer Cell* **5**, 465–475 (2004).
- Mendes-da-Silva, P., Moreira, A., Duro-da-Costa, J., Matias, D. & Monteiro, C. Frequent loss of heterozygosity on chromosome 5 in non-small cell lung carcinoma. *Mol. Pathol.* **53**, 184–187 (2000).
- Korgaonkar, C. *et al.* Nucleophosmin (B23) targets Arf to nucleoli and inhibits its function. *Mol. Cell. Biol.* **25**, 1258–1271 (2005).
- Li, J., Zhang, X., Sejas, D. P., Bagby, G. C. & Pang, Q. Hypoxia-induced Nucleophosmin protects cell death through inhibition of p53. *J. Biol. Chem.* **279**, 41275–41279 (2004).
- Lanni, J. S. & Jacks, T. Characterization of the p53-dependent postmitotic checkpoint following spindle disruption. *Mol. Cell. Biol.* **18**, 1055–1064 (1998).
- Khan, S. H. & Whal, G. M. p53 and pRb prevent rereplication in response to microtubule inhibitors by mediating a reversible G1 arrest. *Cancer Res.* **58**, 369–401 (1998).
- Stewart, Z. A., Leach, S. D. & Pietsenpol, J. A. p21(Waf1/Cip1) inhibition of cyclin E/Cdk2 activity prevents endoreduplication after mitotic spindle disruption. *Mol. Cell. Biol.* **19**, 205–215 (1999).
- Spike, B. T. *et al.* The Rb tumour suppressor is required for stress erythropoiesis. *EMBO J.* **23**, 4319–4329 (2004).
- Lecher-Gagg, A. *et al.* Pescadillo is essential for nucleolar assembly, ribosome biogenesis, and mammalian cell proliferation. *J. Biol. Chem.* **277**, 45347–45355 (2002).
- Bunting, M., Bernstein, K. E., Greer, J. M., Capecchi, M. R. & Thomas, K. R. Targeting genes for self-excision in the germ line. *Genes Dev.* **13**, 1524–1528 (1999).
- Bardeesy, N. *et al.* Loss of the Lkb1 tumour suppressor provokes intestinal polyposis but resistance to transformation. *Nature* **419**, 162–167 (2002).
- Chung, Y. J. *et al.* A whole-genome mouse BAC microarray with 1-Mb resolution for analysis of DNA copy number changes by array comparative genomic hybridization. *Genome Res.* **14**, 188–196 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to M. Leversha and A. Viale at the Molecular Cytogenetics and Genomics Core Facilities and J. Teruya-Feldstein for haematopathology consultation. We thank A. Walz, T. Merghoub, D. Ruggero, E. Hernandez and C. Cordon-Cardo for help and advice; M. Capecchi, N. Bardeesy, R.A. Depinho and I. Zhon for reagents; T. Maeda, L. Montanaro, R. Hobbs, J. Clohessy, L. DiSantis, L. Longo and the other members of the P.P.P. laboratory for discussion, critical reading of the manuscript and support. This work was funded by National Institutes of Health grants to P.P.P.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to P.P.P. (p-pandolfi@ski.mskcc.org).

LETTERS

Structural mechanism for sterol sensing and transport by OSBP-related proteins

Young Jun Im^{1,3}, Sumana Raychaudhuri², William A. Prinz² & James H. Hurley¹

The oxysterol-binding-protein (OSBP)-related proteins (ORPs) are conserved from yeast to humans^{1,2}, and are implicated in the regulation of sterol homeostasis^{3,4} and in signal transduction pathways⁵. Here we report the structure of the full-length yeast ORP Osh4 (also known as Kes1) at 1.5–1.9 Å resolution in complexes with ergosterol, cholesterol, and 7-, 20- and 25-hydroxy-cholesterol. We find that a single sterol molecule binds within a hydrophobic tunnel in a manner consistent with a transport function for ORPs. The entrance is blocked by a flexible amino-terminal lid and surrounded by basic residues that are critical for Osh4 function. The structure of the open state of a lid-truncated form of Osh4 was determined at 2.5 Å resolution. Structural analysis and limited proteolysis show that sterol binding closes the lid and stabilizes a conformation favouring transport across aqueous barriers and signal transmission. The structure of Osh4 in the absence of ligand exposes potential phospholipid-binding sites that are positioned for membrane docking and sterol exchange. On the basis of these observations, we propose a model in which sterol and membrane binding promote reciprocal conformational changes that facilitate a sterol transfer and signalling cycle.

OSBP was first discovered as a cytosolic receptor for oxysterols^{6,7} that downregulate cholesterol synthesis⁸. The cloning of OSBP (ref. 1) led to the identification of a large family of OSBP-related proteins (the ORPs) with seven members in *Saccharomyces cerevisiae* and twelve in *Homo sapiens* (ref. 2). ORPs are essential for life in eukaryotes. The deletion of all seven ORPs leads to severe defects in sterol and lipid distribution and endocytosis in yeast^{3,4}, and OSBP is essential for embryonic development in mice (M. Brown, personal communication). All ORPs contain a core OSBP-related domain, and many also contain pleckstrin homology domains, transmembrane regions, endoplasmic-reticulum-targeting FFAT (two phenylalanines in an acidic tract) motifs, GOLD (Golgi dynamics) domains and/or ankyrin repeats². These additional domains localize ORPs by binding to phosphoinositides⁹, the endoplasmic reticulum protein VAP (vesicle-associated membrane protein (VAMP)-associated protein)¹⁰ and other targeting signals. The localization of ORPs is dynamic. Oxysterol binding changes the subcellular localization of certain ORPs from the cytosol to the Golgi or endoplasmic reticulum^{9,11}. ORPs can bind lipids other than oxysterols, including phosphoinositides and phosphatidic acid^{12,13}. OSBP is a cholesterol-sensing regulator of two protein phosphatases—a PTPPBS (protein tyrosine phosphatase PC12, Br7, Sl) family member and the serine/threonine phosphatase PP2A (protein phosphatase 2A)⁵.

The structure of full-length Osh4 was determined by multiple isomorphous replacement and refined to an R_{free} value of 23% at 1.5 Å resolution (see Methods; Supplementary Fig. 1; Supplementary Table 1). Osh4 is built around a central antiparallel β -sheet of 19

strands (residues 115–293) that form a near-complete β -barrel (Fig. 1a, Supplementary Fig. 2). The strands of the barrel are pitched at $\sim 45^\circ$ to its axis. The barrel has structural similarity to the large β -barrels of various bacterial outer-membrane transporters¹⁴ as scored by DALI (ref. 15) (Supplementary Fig. 3). There is no pleckstrin homology domain within the OSBP-related domain¹³. A tunnel with mainly hydrophobic walls runs through the centre of the barrel in a location that corresponds to the pore in bacterial outer-membrane transporters (Fig. 1b). As a soluble β -barrel protein with a hydrophilic exterior and a hydrophobic tunnel in the barrel centre, Osh4 resembles an inside-out porin. Residues 1–29 form a lid that covers the tunnel opening (Fig. 1c, d). The N-terminal domain following the lid (residues 30–117) consists of a two-stranded β -sheet and three α -helices that form a 50-Å-long antiparallel bundle, which runs the entire length of the barrel. The portion of the bundle distal to the lid fills the centre of the barrel, forming a plug at the far end of the tunnel. The proximal portion of the bundle forms one wall of the tunnel by replacing the missing strands of the barrel. There is a large carboxy-terminal region (residues 308–434) following the barrel. The exterior surface around the lid of the tunnel contains six highly conserved basic residues (Fig. 1c; Supplementary Fig. 4).

Sterols bind to Osh4 within the central tunnel of the β -barrel in a head-down orientation. The 3-hydroxyl group of the sterol is buried at the bottom of the Osh4 tunnel, and the side chain touches the inner surface of the lid (Fig. 1e). Although Osh4 has a novel fold, the burial of its ligands in a central hydrophobic tunnel is reminiscent of the structures of other lipid binding and transport proteins. The closest analogy is to the steroidogenic acute regulatory protein (StAR) transport (START) domain proteins MLN64 (ref. 16) and StarD4 (ref. 17) that bind cholesterol, the START-domain-containing protein phosphatidylcholine-transfer protein¹⁸, and the mammalian phosphatidylinositol-transfer proteins^{19–21}. In these structures, the ligands are completely sequestered from solution. The yeast Sec14 phospholipid-transfer protein²² (and its relatives^{23,24}), the GM2 activator protein²⁵, the cholesterol-binding proteins NPC2 (ref. 26) and Scp2 (refs 27, 28), and the glycolipid transfer protein²⁹ have similar ligand-binding tunnels, although some of these are open at one end^{22,25,29}. The structural analogy to other lipid transporters suggested to us that Osh4 might be an ergosterol transporter, and that other ORPs might also transport sterols or other lipids.

The most striking aspects of oxysterol recognition by Osh4 are the absence of direct hydrogen bonds between sterol hydroxyl groups and the side chains of conserved amino acids within Osh4, and the prominence of water-mediated interactions. The 3-hydroxyl group of cholesterol, ergosterol and oxysterols binds to two water molecules and to the side chain of Gln 96 within Osh4. This Gln is part of a

¹Laboratory of Molecular Biology and ²Laboratory of Cell Biochemistry and Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, US Department of Health and Human Services, Bethesda, Maryland 20892, USA. ³Department of Life Science, Gwangju Institute of Science and Technology, Oryongdong 1, Bukgu, Gwangju City 500-712, South Korea.

hydrated cluster of polar side chains at the bottom of the Osh4 tunnel (Fig. 1e, g). Trp 46, Tyr 97, Asn 165 and Gln 181 comprise the remainder of the polar cluster at the tunnel bottom, and form water-mediated interactions with the 3-hydroxyl group. The 20- and 25-hydroxyl groups of the respective oxysterols interact with relatively ordered water molecules but not directly with the protein. The 7-hydroxyl group of 7-hydroxycholesterol has no apparent hydrogen bonds to either protein or waters. The 25-hydroxyl group is linked through two water molecules to the conserved Lys-109 side chain, but other oxysterol hydroxyl groups are linked to less-well-ordered water molecules. The lack of direct protein interactions with oxysterol hydroxyl groups in Osh4 contrasts sharply to the specific recognition of 24(S),25-epoxycholesterol by a His–Trp pair in the liver X receptor³⁰, a specific effector molecule for a subset of oxysterols.

The absence of direct interactions between the hydroxyl groups and the protein is difficult to reconcile with the concept that Osh4 is a specific effector of oxysterol signalling. We examined the relative affinities of Osh4 for cholesterol and oxysterols. Cholesterol binds to Osh4 with a K_d of 300 nM (Fig. 2a), as compared to a K_d of 55 nM for the oxysterol with the highest affinity among those tested, 25-hydroxycholesterol (Supplementary Fig. 5). Given the modest difference in affinities between these two molecules, and that cholesterol is

far more abundant than oxysterols in mammalian cells, we propose that cholesterol is a physiological ligand for ORPs. We will report elsewhere the results of a study that show Osh4 transports ergosterol, the yeast counterpart of cholesterol (S.R., Y.J.I., J.H.H. and W.A.P., manuscript in preparation).

In the bound conformation of Osh4, the sterol ligands within the complex are inaccessible from the outside. Therefore, a conformational change in Osh4 is required for the uptake and release of sterols. The lid has some of the highest *B* values in the structure (Fig. 3b), which suggests flexibility within this region. After limited proteolysis by trypsin in the absence of ligand, Osh4 is rapidly converted to a stable fragment beginning at residue 28 (Fig. 3a), which corresponds to the end of the lid region (Fig. 3b). Binding of 25-hydroxycholesterol stabilizes this region against proteolysis (Fig. 3a). Sterol ligands stabilize the closed conformation of the lid through direct van der Waals interactions with Trp 10, Phe 13 and the highly conserved residues Leu 24 and Leu 27 (Fig. 1f, g). We were unable to crystallize full-length Osh4 in the absence of ligand, which we attribute to the flexibility of the lid region in the unliganded state. We engineered a ‘lidless’ Osh4 in which residues 1–29 were deleted and the flexible surface loop 236–240 was replaced by an ectopic dipeptide sequence. Mutation of loop 236–240 does not affect cholesterol binding (data not shown). We determined the structure

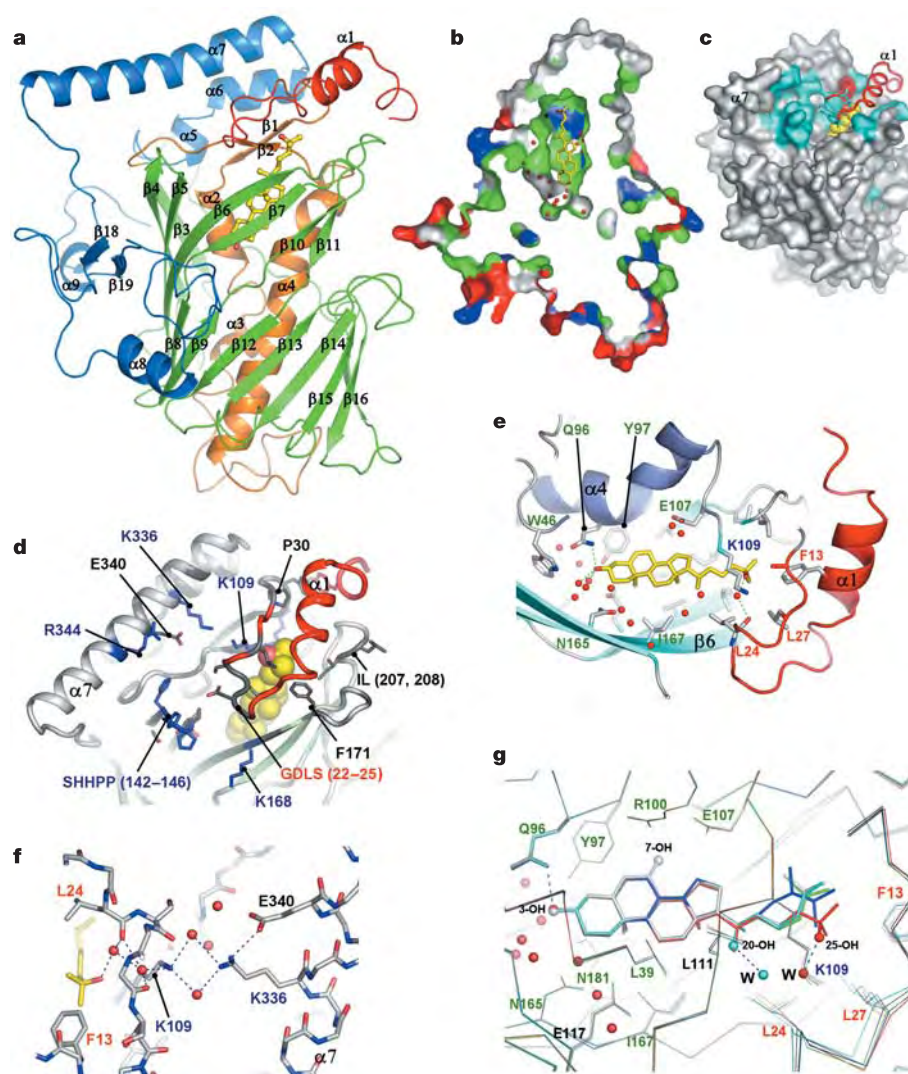


Figure 1 | Structure of Osh4. **a**, The overall structure of Osh4. The N-terminal lid (residues 1–29) is shown in red, the central helices (30–116) in orange, the β -barrel (117–307) in green and the C-terminal subdomain (308–434) in cyan. The yellow structure is 25-hydroxycholesterol. **b**, The surface of Osh4. Basic residues are shown in blue, acidic residues in red, hydrophobic residues in green and neutral polar residues in white. Water molecules are shown as red spheres. The yellow structure is 25-hydroxycholesterol. **c**, The N-terminal lid of Osh4. Strictly conserved residues are coloured in cyan. The N-terminal lid is shown as ribbons and coloured in red. The yellow structure is 25-hydroxycholesterol. **d**, Tunnel-entrance residues of Osh4. Solvent-accessible conserved residues are clustered around the tunnel entrance. Basic residues are shown in blue, residues within the lid in red, hydrophobic residues (not in the lid) in black. The yellow structure is 25-hydroxycholesterol. **e**, Hydration and hydrogen bonding within Osh4. Residues that bind 25-hydroxycholesterol are coloured in green. Water molecules are shown as red spheres. Hydrogen bonds are shown as dashed lines. The yellow structure is 25-hydroxycholesterol. **f**, Recognition of the 25-hydroxycholesterol 25-hydroxyl group. The ϵ -amino group of Lys 336 has two conformations in all complexed crystal structures, with the left-hand conformation (closest to Lys 109) predominating. Water molecules are shown as red spheres. Hydrogen bonds are shown as dashed lines. The yellow structure is 25-hydroxycholesterol. **g**, Superposition of five sterols in the Osh4 binding site. 7-Hydroxycholesterol is coloured in grey, 20-hydroxycholesterol in cyan, 25-hydroxycholesterol in red, cholesterol in green and ergosterol in blue, with the respective hydroxyl groups labelled. Water molecules are shown as spheres. Hydrogen bonds are shown as dashed lines. W indicates key water molecules with which the 20- and 25-hydroxyl groups of the respective oxysterols interact.

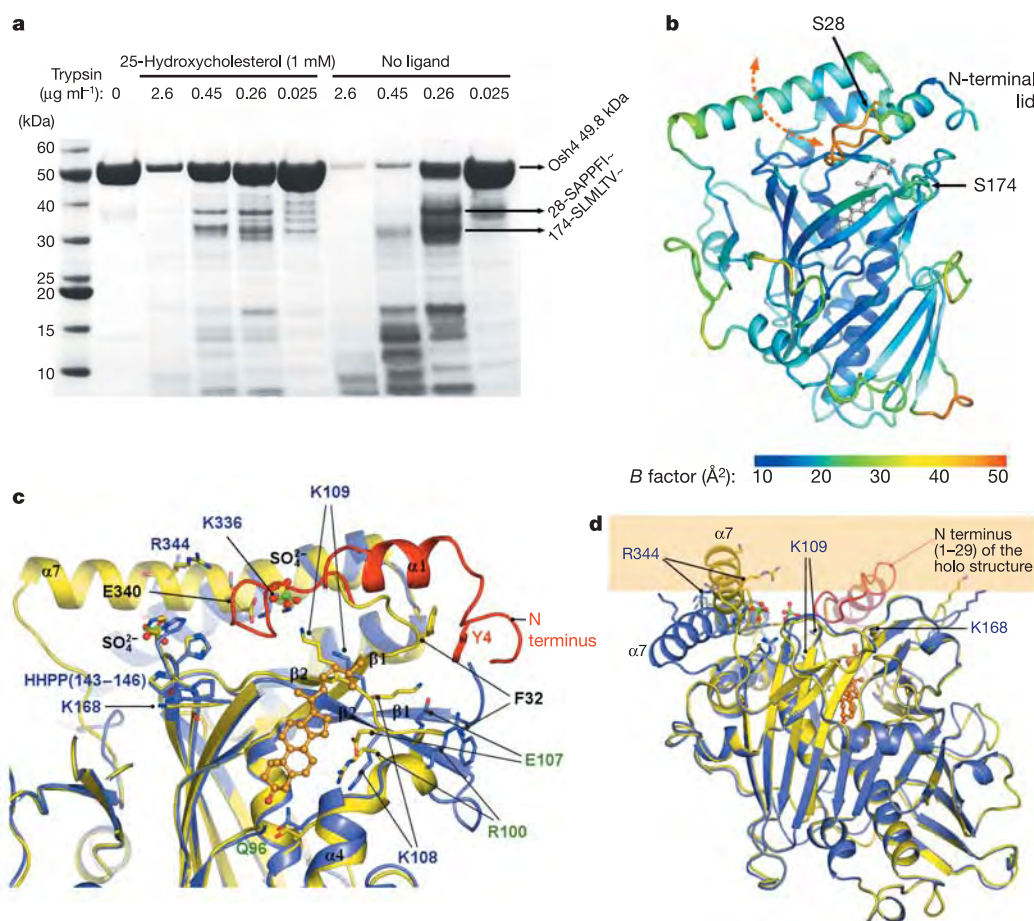
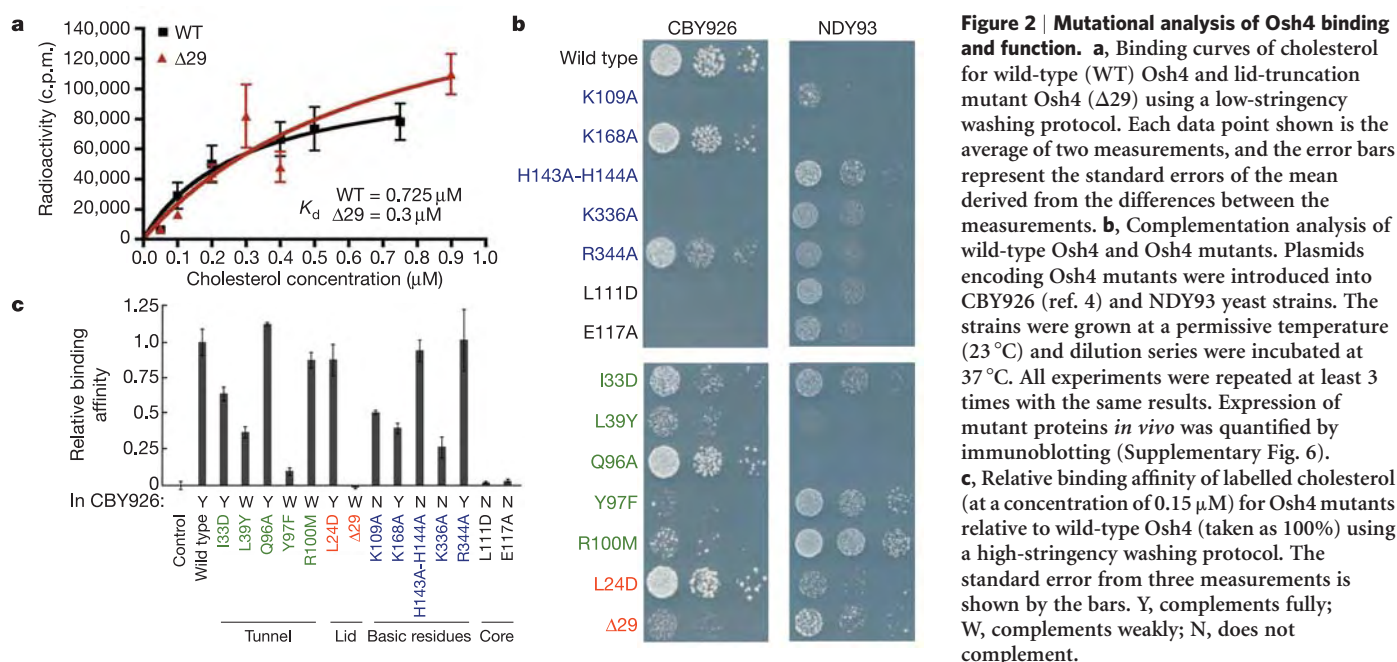


Figure 3 | Conformational changes in Osh4. **a**, Tryptic digestion of Osh4 with increasing concentrations of trypsin. The arrows indicate the N-terminal amino acid sequences of the bands. **b**, The ribbon model is coloured by the average *B* factor of the residues to show their relative mobility. Trypsin cleavage sites are indicated by arrows. The white structure is 25-hydroxycholesterol. **c**, Apo (blue, residues 30–434) and cholesterol-

complex (red for the lid region, residues 2–29; yellow for elsewhere, residues 30–434) structures were superimposed. The sulphate ions bound to the apo structure in positions that membrane phospholipid phosphates would bind are shown. The orange structure is cholesterol. **d**, Superposition of apo and cholesterol complex (holo) structures with the location of the membrane indicated by yellow shading. The orange structure is cholesterol.

of this lidless form of Osh4 in its unliganded conformation at 2.5 Å resolution, which we refer to below as the 'apo' structure (Fig. 3c, d).

The apo structure undergoes a dramatic conformational change (2.9 Å r.m.s. for all common C α positions) compared with the complexed structure. This conformational change is triggered when the ligand–lid interactions are lost. The movement of the lid away from the rest of the protein unlocks helix α 7 and basic-residue-containing loops from their complexed conformations. In the apo structure, the tunnel is open and there is no longer any barrier to ligand exchange. Helix α 7 pivots about its N terminus such that its C terminus moves 15 Å. The β 1 strand and several loop regions near the tunnel opening move by up to 7 Å. The C α of the conserved basic residue Lys 109 moves 6 Å in this conformational change. These shifts result in the reorganization of the entire conserved basic cluster at the tunnel entrance (Fig. 3c). The apo conformation presents putative phosphate-binding sites and a flattened unobstructed surface surrounding the tunnel opening, which are not present in the bound conformation (Fig. 3d).

We tested whether the conserved basic residues at the tunnel entrance of Osh4, as well as structural core, lid and sterol-binding residues, were important for biological function by assessing their ability to complement a yeast strain in which all seven ORPs have either been deleted, or are present as a temperature-sensitive allele³. Lys 109, the His 143–His 144 pair and Lys 336 are essential for function¹³, whereas two basic residues that lie further from the tunnel entrance (Lys 168 and Arg 344) are not (Fig. 2b, Supplementary Fig. 6). None of these residues are in direct contact with cholesterol, and their mutation only modestly impairs cholesterol binding (Fig. 2c). In contrast, some mutations within the tunnel (such as Y97F) and structure core mutations near the tunnel (such as L111D and E117A) abrogate both cholesterol binding and biological function (Fig. 2a, c). K109A, K168A and K336A mutations all lead to sharp reductions in basal cholesterol transport by Osh4 (S.R. *et al.*, manuscript in preparation), while H143A–H144A shows a smaller reduction; hence, the major biochemical consequence of mutation within the basic cluster is a consistent defect in sterol transport.

In the complexes, the ϵ -amino groups of Lys 109 and Lys 336 are $\sim$ 3.6 Å away from each other, and lie in an unfavourable semi-buried environment (Fig. 1f). In the apo structure, they move 8.7 Å apart.

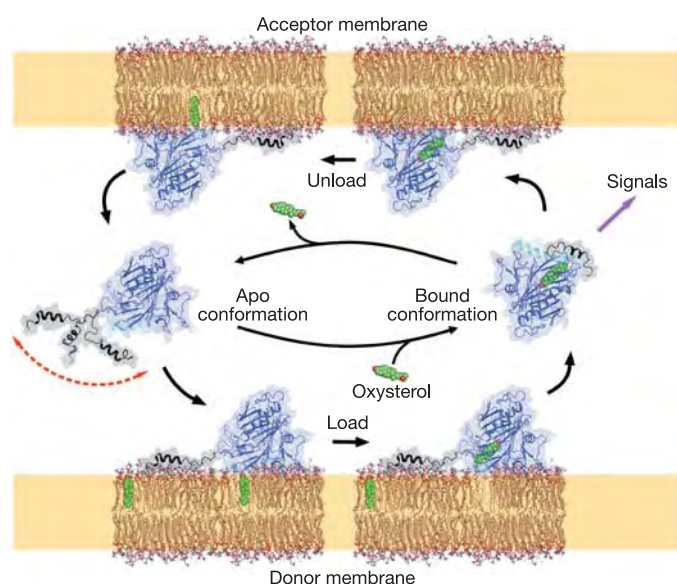


Figure 4 | Mechanism of sterol transfer. The proposed cholesterol transfer cycle is depicted. Oxysterols are shown binding in the cytosol rather than the membrane (as for cholesterol) because of their higher solubility. Cholesterol and oxysterol-dependent signals are shown as occurring through the bound conformation in the cytosol as found for OSBP (ref. 5).

Two ordered sulphate ions are bound in the apo structure: one is bound between the new positions of Lys 109 and Lys 336, and the other near the conserved His pair (Fig. 3c). We believe these sulphate ions mimic the roles of phospholipid phosphate groups because they bind in the position expected for the plane of the membrane surface during sterol exchange. The unfavourable geometry of the Lys 109–Lys 336 pair in the complex suggests that the closed structure is in a destabilized 'tense' state. We propose that the open state is selectively promoted by the availability of phosphate groups at the membrane surface, which bind to these Lys residues only in the open state. Thus, these residues appear to function as an electrostatic spring-loaded lock that controls access to the tunnel. On the basis of these observations, we propose a sterol transfer and signalling cycle (Fig. 4).

OSBP is a cholesterol sensor that regulates phosphatase complexes⁵. 25-hydroxycholesterol antagonizes cholesterol activation. We find that 25-hydroxycholesterol and cholesterol bind to the same site within Osh4 and induce similar gross structural changes. However, 25-hydroxycholesterol is unique among oxysterols in that it forms a water-mediated hydrogen bond with the critical Lys 109 residue. Differences in packing between the side chains of different sterols and the surrounding side chains of Ile 17, Leu 24 and Leu 27 within Osh4 lead to shifts in the lid main-chain position by up to 1.5 Å when compared with the cholesterol complex (Supplementary Fig. 7). Within the oxysterol complexes, the 25-hydroxycholesterol structure shows the largest differences from the cholesterol complex. The possible role that these subtle structural changes have in explaining 25-hydroxycholesterol/cholesterol antagonism warrants further study.

The lid region and the basic cluster at the tunnel opening are conserved in all ORPs, whereas parts of the cholesterol-binding site are not well conserved, suggesting that many ORPs might transport non-sterol ligands. The conservation of the lid and basic cluster lead us to believe that the transport cycle will be maintained throughout this protein family. Similarly, cholesterol-dependent signalling will require the same apparatus as cholesterol transfer in order to facilitate the uptake of cholesterol at membranes.

METHODS

Protein expression and purification. DNA encoding Osh4 residues 2–434 was amplified by polymerase chain reaction and subcloned into the *Bam*HI and *Xho*I sites of a pGEX-4T vector modified to contain a tobacco etch virus (TEV) protease recognition site. The glutathione S-transferase (GST) fusion protein was expressed in *Escherichia coli* BL21(DE3) cells overnight at 30 °C. The cellular lysate was applied to glutathione sepharose, and the GST removed by digestion with TEV protease in the column. Osh4 was eluted with 500 mM NaCl and 50 mM NaH₂PO₄ pH 7.5, concentrated and then purified on a Superdex 200 column (Pharmacia) in 20 mM Tris-HCl pH 8.0, 100 mM NaCl.

Crystallization. 25-Hydroxycholesterol was prepared in ethanol and added at a final concentration of 1 mM to 15 mg ml^{−1} of protein. Crystals were grown by vapour-diffusion at 25 °C above a reservoir of 100 mM 2-N-morpholinoethanesulphonic acid (MES)-NaOH pH 6.5, 12% (w/v) polyethylene glycol 20,000 over one week. Crystals for complexes with ergosterol, cholesterol, 7-hydroxycholesterol and 20-hydroxycholesterol were similarly grown. Crystals were cryoprotected in reservoir solution supplemented with 20% (v/v) glycerol and flash frozen under nitrogen gas at 95 K. Apo Osh4 crystals were grown in 100 mM MES-NaOH pH 6.0 and 1.6 M ammonium sulphate using micro- and macro-seeding. The apo crystals were cryoprotected in reservoir solution containing 32% (v/v) glycerol. Structure determination is described in the Supplementary Methods.

Tryptic digestion and N-terminal amino acid sequencing. 25-Hydroxycholesterol-bound Osh4 was prepared by adding 25-hydroxycholesterol to a final concentration of 1 mM to a 0.3 mM protein solution. The 25-hydroxycholesterol-bound Osh4 or apo-Osh4 proteins (30 μ g) were digested with increasing trypsin concentrations at 37 °C for 2 h. Fragments of the Osh4 protein prepared by limited proteolysis with trypsin were separated by SDS–polyacrylamide gel electrophoresis, transferred to polyvinylidene fluoride membrane, and stained with Coomassie blue R-250. Bands were cut out and the N-terminal amino acids were sequenced using a 492cLC Protein Sequencer (Applied Biosystems).

Ligand-binding assay. For the analysis of mutant Osh4 binding at a single concentration (Fig. 2a), [³H]cholesterol and unlabelled cholesterol were mixed

at a ratio of 1:9 to a final concentration of 10 μ M in ethanol. Each reaction was carried out at 22 °C in a final volume of 100 μ l of 20 mM Tris-HCl pH 8.5, 30 mM NaCl and 100 pmol Osh4. Cholesterol (15 pmol) was added to the protein solution and incubated for 1 h. HiTrap Q resin (20 μ l of 70% (w/v); Amersham) pre-equilibrated with buffer A (20 mM Tris-HCl pH 8.5) was added and incubated for 5 min. The tubes were centrifuged at 13,000g for 1 min and washed four times with 1 ml of buffer A. The protein was eluted from the resin with 700 μ l of buffer A containing 1 M NaCl. After centrifuging at 13,000g for 10 min, 200 μ l of supernatant was taken and the radioactivity measured in a liquid scintillation counter. In the low-stringency washing protocol (Fig. 2b), radiolabelled cholesterol (40 pmol) was dried under nitrogen and resuspended in 100 μ l of 20 mM Tris-HCl pH 8.5, 30 mM NaCl, 0.05% Triton-X100. Purified Osh4 protein was added to a final concentration of 200 nM and incubated at 30 °C. After incubation for 1 h, 15 μ l of HiTrap Q resin was added and the samples were vortexed and left at 22 °C for 5 min. The resin was pelleted by spinning at top speed in a microcentrifuge for 5 min. After three washes with 20 mM Tris pH 8.5, the protein was eluted from the resin with 500 μ l of 20 mM Tris-HCl pH 8.5, 1 M NaCl. After centrifugation, the radiolabelled cholesterol in the supernatant was determined by scintillation counting. To determine non-specific binding, 500 μ M of unlabelled cholesterol was included in the incubation.

Yeast strains and complementation analysis. Complementation analysis was performed by introducing plasmids encoding mutant Osh4 proteins into either the CBY926 (ref. 4) or NDY63 (*MATa sec14-3 osh4::kanMX4 ura3-1 his3-11,-15 leu2-5,-112*) yeast strains. CBY926 contains a temperature-sensitive allele of *OSH4* (*osh4-1*) and deletions of the other six *OSH* genes. NDY93 is a *sec14-1 Δ osh4* strain that cannot grow at 37 °C if it contains a plasmid encoding a functional Osh4 protein; complementing plasmids prevent NDY93 from growing at 37 °C (ref. 13). To assess complementation, strains containing the various plasmids were plated on synthetic complete dextrose medium minus uracil and grown at 37 °C.

Received 16 March; accepted 9 June 2005.

- Dawson, P. A., Ridgway, N. D., Slaughter, C. A., Brown, M. S. & Goldstein, J. L. cDNA cloning and expression of oxysterol-binding protein, an oligomer with a potential leucine zipper. *J. Biol. Chem.* **264**, 16798–16803 (1989).
- Olkonen, V. M. & Levine, T. P. Oxysterol binding proteins: in more than one place at one time? *Biochem. Cell Biol.* **82**, 87–98 (2004).
- Beh, C. T., Cool, L., Phillips, J. & Rine, J. Overlapping functions of the yeast oxysterol-binding protein homologues. *Genetics* **157**, 1117–1140 (2001).
- Beh, C. T. & Rine, J. A role for yeast oxysterol-binding protein homologs in endocytosis and in the maintenance of intracellular sterol-lipid distribution. *J. Cell Sci.* **117**, 2983–2996 (2004).
- Wang, P., Weng, J. & Anderson, R. G. W. OSBP is a cholesterol-regulated scaffolding protein in control of ERK1/2 activation. *Science* **307**, 1472–1476 (2005).
- Kandutsch, A. A., Chen, H. W. & Heiniger, H. J. Biological activity of some oxygenated sterols. *Science* **201**, 498–501 (1978).
- Taylor, F. R., Saucier, S. E., Shown, E. P., Parish, E. J. & Kandutsch, A. A. Correlation between oxysterol binding to a cytosolic binding-protein and potency in the repression of hydroxymethylglutaryl coenzyme-A reductase. *J. Biol. Chem.* **259**, 12382–12387 (1984).
- Schroepfer, G. J. Jr Oxysterols: Modulators of cholesterol metabolism and other processes. *Physiol. Rev.* **80**, 361–554 (2000).
- Levine, T. P. & Munro, S. The pleckstrin homology domain of oxysterol-binding protein recognises a determinant specific to Golgi membranes. *Curr. Biol.* **8**, 729–739 (1998).
- Loewen, C. J. R., Roy, A. & Levine, T. P. A conserved ER targeting motif in three families of lipid binding proteins and in Opi1p binds VAP. *EMBO J.* **22**, 2025–2035 (2003).
- Ridgway, N. D., Dawson, P. A., Ho, Y. K., Brown, M. S. & Goldstein, J. L. Translocation of oxysterol binding-protein to Golgi-apparatus triggered by ligand-binding. *J. Cell Biol.* **116**, 307–319 (1992).
- Xu, Y. Q., Liu, Y. L., Ridgway, N. D. & McMaster, C. R. Novel members of the human oxysterol-binding protein family bind phospholipids and regulate vesicle transport. *J. Biol. Chem.* **276**, 18407–18414 (2001).
- Li, X. M. *et al.* Analysis of oxysterol binding protein homologue Kes1p function in regulation of Sec14p-dependent protein transport from the yeast Golgi complex. *J. Cell Biol.* **157**, 63–77 (2002).
- Buchanan, S. K. β -Barrel proteins from bacterial outer membranes: structure, function and refolding. *Curr. Opin. Struct. Biol.* **9**, 455–461 (1999).
- Holm, L. & Sander, C. DALI: a network tool for protein-structure comparison. *Trends Biochem. Sci.* **20**, 478–480 (1995).
- Tsujiishi, Y. & Hurley, J. H. Structure and lipid transport mechanism of a StAR-related domain. *Nature Struct. Biol.* **7**, 408–414 (2000).
- Romanowski, M. J., Soccio, R. E., Breslow, J. L. & Burley, S. K. Crystal structure of the *Mus musculus* cholesterol-regulated START protein 4 (StarD4) containing a StAR-related lipid transfer domain. *Proc. Natl Acad. Sci. USA* **99**, 6949–6954 (2002).
- Roderick, S. L. *et al.* Structure of human phosphatidylcholine transfer protein in complex with its ligand. *Nature Struct. Biol.* **9**, 507–511 (2002).
- Tilley, S. J. *et al.* Structure-function analysis of phosphatidylinositol transfer protein alpha bound to human phosphatidylinositol. *Structure* **12**, 317–326 (2004).
- Schouten, A. *et al.* Structure of apo-phosphatidylinositol transfer protein alpha provides insight into membrane association. *EMBO J.* **21**, 2117–2121 (2002).
- Yoder, M. D. *et al.* Structure of a multifunctional protein. Mammalian phosphatidylinositol transfer protein complexed with phosphatidylcholine. *J. Biol. Chem.* **276**, 9246–9252 (2001).
- Sha, B. D., Phillips, S. E., Bankaitis, V. A. & Luo, M. Crystal structure of the *Saccharomyces cerevisiae* phosphatidylinositol-transfer protein. *Nature* **391**, 506–510 (1998).
- Meier, R., Tomizaki, T., Schulze-Bries, C., Baumann, U. & Stocker, A. The molecular basis of vitamin E retention: structure of human alpha-tocopherol transfer protein. *J. Mol. Biol.* **331**, 725–734 (2003).
- Min, K. C., Kovall, R. A. & Hendrickson, W. A. Crystal structure of human alpha-tocopherol transfer protein bound to its ligand: implications for ataxia with vitamin E deficiency. *Proc. Natl Acad. Sci. USA* **100**, 14713–14718 (2003).
- Wright, C. S., Li, S. C. & Rastinejad, F. Crystal structure of human GM2-activator protein with a novel β -cup topology. *J. Mol. Biol.* **304**, 411–422 (2000).
- Friedland, N., Liou, H. L., Lobel, P. & Stock, A. M. Structure of a cholesterol-binding protein deficient in Niemann-Pick type C2 disease. *Proc. Natl Acad. Sci. USA* **100**, 2512–2517 (2003).
- Choinowski, T., Hauser, H. & Piontek, K. Structure of sterol carrier protein 2 at 1.8 Å resolution reveals a hydrophobic tunnel suitable for lipid binding. *Biochemistry* **39**, 1897–1902 (2000).
- Lopez-Garcia, F. *et al.* NMR structure of the sterol carrier protein-2: implications for the biological role. *J. Mol. Biol.* **295**, 595–603 (2000).
- Malinin, L., Malakhova, M. L., Teplov, A., Brown, R. E. & Patel, D. J. Structural basis for glycosphingolipid transfer specificity. *Nature* **430**, 1048–1053 (2004).
- Williams, S. *et al.* X-ray crystal structure of the liver X receptor β ligand binding domain: regulation by a histidine-tryptophan switch. *J. Biol. Chem.* **278**, 27138–27143 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank T. Levine, V. Bankaitis and M. Brown for discussions and sharing unpublished data, N. DeAngelis for technical assistance, R. Craigie for assistance with protein sequencing, J. Kim for advice on the early stages of this project, G. Miller and H. Shi for collecting synchrotron data, C. Beh and R. Scheckman for yeast strains, F. Dyda for maintaining the home X-ray facility, and the staff of beamline X25, National Synchrotron Light Source, Brookhaven National Laboratory and of SER-CAT, Advanced Photon Source, Argonne National Laboratory for assistance with data collection. This research was supported by the intramural program of the NIDDK. Y.J.I. thanks S. H. Eom for mentoring and support. Y.J.I. was partly supported by the Korea Science and Engineering Foundation. Research carried out at the National Synchrotron Light Source is supported by the US Department of Energy, Division of Materials Sciences and Division of Chemical Sciences. Use of the Advanced Photon Source was supported by the US Department of Energy, Basic Energy Sciences, Office of Science.

Author Information Coordinates have been deposited with the Protein Data Bank with accession numbers 1ZHT, 1ZHW, 1ZHX, 1ZHY, 1ZHZ and 1Z17. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.H.H. (hurley@helix.nih.gov).

Observation of a dewetting transition in the collapse of the melittin tetramer

Pu Liu¹, Xuhui Huang¹, Ruhong Zhou^{1,2} & B. J. Berne^{1,2}

Marked hydration changes occur during the self-assembly of the melittin protein tetramer in water. Hydrophobicity induces a drying transition in the gap between simple sufficiently large (more than 1 nm²) strongly hydrophobic surfaces as they approach each other^{1–6}, resulting in the subsequent collapse of the system, as well as a depletion of water next to single surfaces^{7–10}. Here we investigate whether the hydrophobic induced collapse of multidomain proteins or the formation of protein oligomers exhibits a similar drying transition. We performed computer simulations to study the collapse of the tetramer of melittin in water, and observed a marked water drying transition inside a nanoscale channel of the tetramer (with a channel size of up to two or three water-molecule diameters). This transition, although occurring on a microscopic length scale, is analogous to a first-order phase transition from liquid to vapour. We find that this drying is very sensitive to single mutations of the three isoleucines to less hydrophobic residues and that such mutations in the right locations can switch the channel from being dry to being wet. Thus, quite subtle changes in hydrophobic surface topology can profoundly influence the drying transition. We show that, even in the presence of the polar protein backbone, sufficiently hydrophobic protein surfaces can induce a liquid–vapour transition providing an enormous driving force towards further collapse. This behaviour was unexpected because of the absence of drying in the collapse of the multidomain protein 2,3-dihydroxybiphenyl dioxygenase (BphC).

Proteins exert electrostatic forces on water and are therefore much more complicated than purely hydrophobic systems. We have seen previously that because of these electrostatic forces there was no strong drying transition in the collapse of the BphC enzyme¹¹, a two-domain protein. These findings showed that realistic protein–water forces have a significant function in hydrophobic collapse, and argued against a drying-induced reorganization model of protein folding. Others have examined the role of water during protein folding, finding that water can persist inside the core even after it collapses and can also act as a lubricant for folding^{12,13}, but those studies did not address the drying transition that we discuss here. A coarse-grained simulation of the collapse of a model hydrophobic polymer chain showing that the evaporation (drying) of water in the vicinity of the polymer provides the driving force for the collapse, and that the rate-limiting step is the nucleation of a sufficiently large bubble, caused its authors¹⁴ to speculate, “Our findings would seem also pertinent to the mechanism of biological assembly, such as protein folding, but to demonstrate so with simulation will require an analogous simulation study of a protein-like chain”. Here we study the water drying transition inside the melittin tetramer. Melittin, a 26-residue polypeptide, is a small toxic protein found in honeybee venom, which often self-assembles into a tetramer. Previous simu-

lations of hydration structure around the melittin dimer partly guided our choice of the tetramer for study¹⁵. The two dimers of the melittin tetramer are separated by a certain distance to create a nanoscale water ‘channel’. The enlarged melittin tetramer is then simulated with molecular dynamics. A strong, cooperative water drying transition is found inside the nanoscale channel whose size is large enough to encompass two or three water-molecule diameters, in agreement with previous speculation¹⁴.

For each enlarged melittin tetramer configuration (see Methods), three different types of molecular dynamics simulation were performed¹¹: a ‘dewetting’ simulation starting with wet initial conditions (the channel filled with water molecules, protein atoms constrained); a ‘wetting’ simulation starting with dry initial conditions (no water molecules in the channel, protein atoms constrained); and a normal ‘folding’ simulation. In both the ‘dewetting’ and ‘wetting’ simulations, the protein atoms are constrained with a harmonic potential with a force constant of 10 kJ mol^{−1} Å^{−2}, but water molecules and ions are free to move. For the ‘wetting’ simulation, the water molecules inside the tetramer channel are removed after equilibration. Figure 1a shows snapshots from one such ‘dewetting’ simulation with an initial separation distance D of 5.5 Å. A water vapour bubble developed very quickly, within about 100 ps, which fluctuated for a short period before growing to such a large size that all the water molecules inside the channel were expelled. The entire drying process took only about 300–400 ps in this case. Figure 1b shows a plot of the number of water molecules inside the channel compared with the molecular dynamics time for both ‘dewetting’ and ‘wetting’ simulations. The two water number curves converge after about 400 ps, and at the end no water molecules are observed near the centre of the nanoscale channel. These results clearly indicate that there is a strong water drying transition inside this nanoscale melittin tetramer channel. Other trajectories starting from different initial configurations show similar results (Supplementary Information). Careful simulations with many other separation distances show that the drying transition critical distance D_c for this melittin tetramer system is about 5.5–7.0 Å, which is equivalent to two or three water-molecule diameters.

The kinetics of the collapse of the tetramer from its expanded configuration shows a very fast hydrophobic collapse when the two dimers are within the critical distance. Figure 1c plots a few trajectories of the dimer separation distance against the molecular dynamics time, starting from two different initial distances. The collapse can be very fast (as little as 150 ps) once it has passed through the initial diffusive stage. The collapse will be even faster, within 50 ps, if started from initial ‘dry’ condition. Large driving forces seem to be pushing the tetramers together, as recently observed in simulations of the collapse of paraffin-like plates where the collapse was found to be rate-limited by the nucleation of a vapour bubble in

¹Department of Chemistry, Columbia University, New York, New York 10027, USA. ²Computational Biology Center, IBM Thomas J. Watson Research Center, 1101 Kitchawan Road, Yorktown Heights, New York 10598, USA.

the gap^{5,8,9}. A drying-induced collapse is found in some trajectories (one of them is shown in Fig. 1d), even though other trajectories show that drying and collapse happen at roughly the same time.

What makes this melittin tetramer behave differently from the BphC enzyme, the two-domain protein previously studied¹¹? Even though BphC was chosen because it was thought to contain very hydrophobic domains based on a hydrophobic moment profile analysis, we did not find a water drying transition even at very small domain separations such as $D = 4 \text{ \AA}$ (ref. 11). A quick analysis of the residue composition shows that the melittin tetramer is indeed very special—it has a very strong hydrophobic presence in the inner surface of the tetramer, with three Ile, four Leu, one Trp and two Val residues on each monomer helix. Meanwhile, it has five charged residues (three Lys and two Arg) on each monomer helix pointing away from the centre of the tetramer. It is therefore of great interest to see how sensitive the drying transition is to mutations of some of the key hydrophobic residues. We chose to perform single mutations of the Ile residues, which are the most hydrophobic residues on melittin. The locations of Ile 2, Ile 17 and Ile 20 are shown in Fig. 2a. Three different types of mutation were performed, Ile to Val, Ile to Ala, and Ile to Gly, with decreasing hydrophobicity. All these mutations were single mutations; in other words, only one Ile residue was mutated at any time for each monomer.

Table 1 summarizes the effects of the mutations. As expected, single mutations to Gly have the strongest effect and mutations to Val have the least effect. In addition, for Gly mutations, significantly different results can happen depending on the mutation location: Ile2Gly wets, Ile17Gly dries (dewets), and Ile20Gly fluctuates between wet and dry. In contrast, the effects of mutations to Ala fall somewhere between. Figure 2b shows a plot of the number of water molecules inside the tetramer channel against the molecular dynamics simulation time for all three Ala mutations (see Supplementary Information for more trajectories). The Ile17Ala mutation (green curve) still shows a sharp drying transition, with water density inside the channel decreasing to zero after about 200 ps. However, the Ile2Ala mutation (red curve) shows no drying transition during the entire simulation. The Ile20Ala mutation (blue curve) leads to larger fluctuations in water density and dewets noticeably more slowly (after 600 ps). The most interesting finding from Table 1, however, is that Ile 2 seems to be the most sensitive residue: all three single mutations at that position cause the water drying transition to disappear. Considering that Val is also a very hydrophobic residue, the Ile 2 position is indeed very sensitive to mutations. A closer look shows that residue Ile 2 protrudes from the dimer plane, acting as a 'bump' or 'door' to the tetramer channel (see Fig. 3a). The Ile 20 residue also shows a bump, but not as large as that

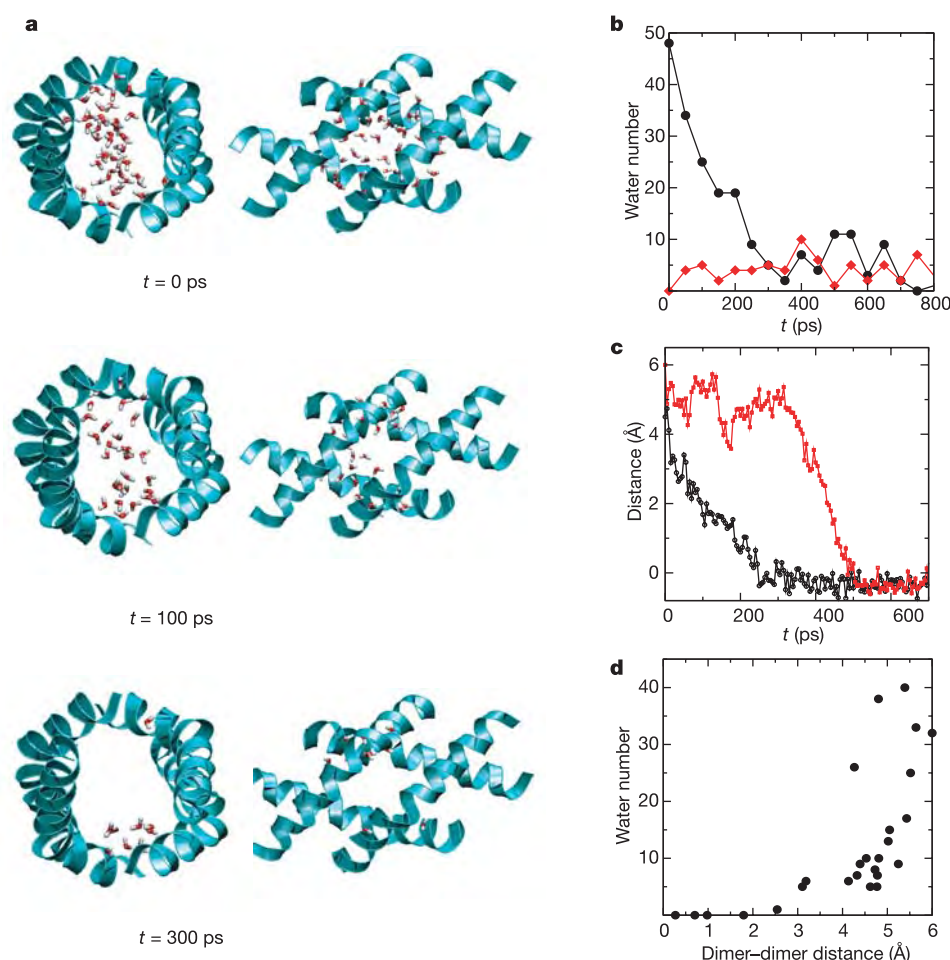


Figure 1 | Dewetting in melittin. **a**, Snapshots of water molecules inside the gap of the melittin tetramer (the protein is shown as ribbons and water as sticks). Only water molecules near the centre of the channel are plotted, which is defined as the region with a spherical radius of $10 \text{ \AA}$ from the centre of the enlarged tetramer. **b**, Plot of the number of water molecules inside the channel against the molecular dynamics time for 'dewetting' (black) and 'wetting' (red) simulations. **c**, The kinetics of the 'folding' simulation

starting from two different initial separations, $D = 6 \text{ \AA}$ (black) and $D = 4.5 \text{ \AA}$ (red). **d**, Plot of the number of water molecules inside the channel against the dimer-dimer distance for one folding trajectory starting at an initial separation of $6 \text{ \AA}$, which indicates a drying-induced collapse (see Supplementary Information for more trajectories showing drying and collapse happening at roughly the same time).

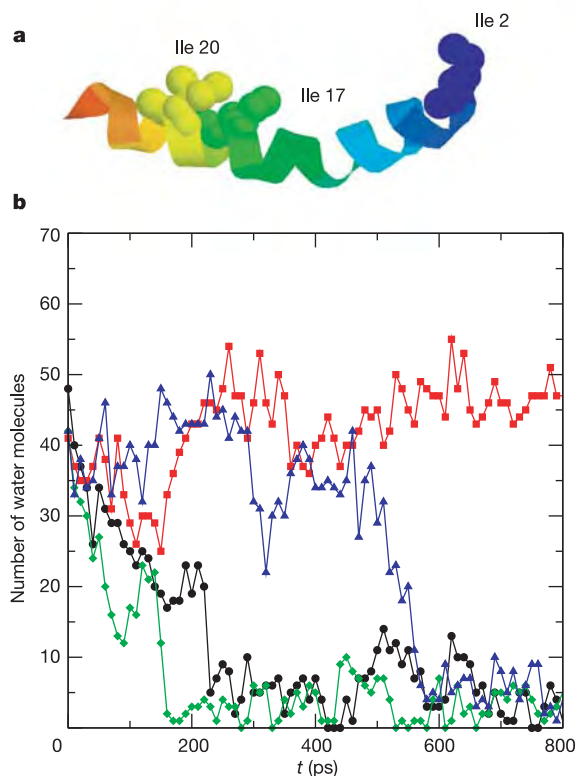


Figure 2 | Dewetting in mutated melittin. **a**, Illustration of three mutation sites, Ile 2, Ile 17 and Ile 20, on one monomer. Ile 2 is sticking out the dimer plane, and Ile 17 is lying on the plane; Ile 20 is somewhat tilted, with a conformation between those of Ile 2 and Ile 17. **b**, The number of water molecules in the tetramer channel with $D = 5.5 \text{ \AA}$ as a function of simulation time. It is clear that the dewetting strength is sensitive to the surface topology of the protein. Black, no mutation; red, Ile2Ala; green, Ile17Ala; blue, Ile20Ala.

for Ile 2, whereas Ile 17 lies on the dimer plane. This special surface topology might be responsible for Ile2's central role in the drying transition in this melittin tetramer.

To improve our understanding of the function of the Ile 2 residue in surface topology, we designed a model system to mimic the shape

Table 1 | Dewetting results from single mutations on three Ile residues of each melittin monomer

Model	Ile 2	Ile 17	Ile 20
No mutation	Dewet	Dewet	Dewet
Mutate to Val	Wet	Dewet	Dewet
Mutate to Ala	Wet	Dewet	Dewet
Mutate to Gly	Wet	Dewet	Fluctuation

characteristics of the melittin tetramer. Figure 3b shows a model of two graphite-like plate systems. Two argon clusters were grown to mimic the surface bumps on the graphite surface. Each graphite-like plate has 170 carbon atoms with C–C Lennard–Jones interaction reduced to $\epsilon_{cc} = 0.0213 \text{ kcal mol}^{-1}$ (for comparison, ϵ_{cc} for graphite is $0.0860 \text{ kcal mol}^{-1}$). This results in a plate with a radius of about $14 \text{ \AA}$. Each argon cluster has 13 argon atoms (with $\epsilon = 0.00656 \text{ kcal mol}^{-1}$) and the distance from each argon atom to the nearest surface carbon atom is about $2.2 \text{ \AA}$. For this graphite-like plate system, the critical distance is $4.5\text{--}6.5 \text{ \AA}$ (measured from one plate surface to the other plate surface) without the bumps, and it increases to about $6.5\text{--}8.5 \text{ \AA}$ with the bumps. Thus, these argon cluster bumps have increased the transition critical distance by about $2 \text{ \AA}$. These surface bumps disrupt the hydrogen-bond network of water between the plates, thus making the wet configurations less favourable. This disruption is probably even more extensive for the tetramer channel, given its greater surface irregularity. The water hydrogen-bond network is energetically more critical in the tetramer channel case than the two-plate case, because the channel is one-dimension-like whereas the inter-plate region is two-dimensional.

There seem to be two reasons why the melittin tetramer channel shows a drying transition while the two-domain protein BphC does not. The melittin channel is like a tube, whereas the inter-domain region in the two-domain protein is a slab. It is less costly in terms of free energy to disrupt the hydrogen bonds in a tube-like channel. Moreover, the special surface topology springing from the Ile residues destabilizes the wetting in the channel. Even though there has been no direct experimental evidence yet for this nanoscale dewetting transition in real proteins, two recent experiments have shown partial dewetting for paraffin-like systems^{1,16,17}: one for $^2\text{H}_2\text{O}$ in contact with deuterated polystyrene in a neutron reflectivity experiment¹⁶, and the other using water in contact with paraffin in an X-ray reflectivity experiment¹⁷. Partial dewetting was found, with a water density about 10% lower than the bulk near the interface. It would be of interest to devise experiments to test our predictions above for melittin tetramers.

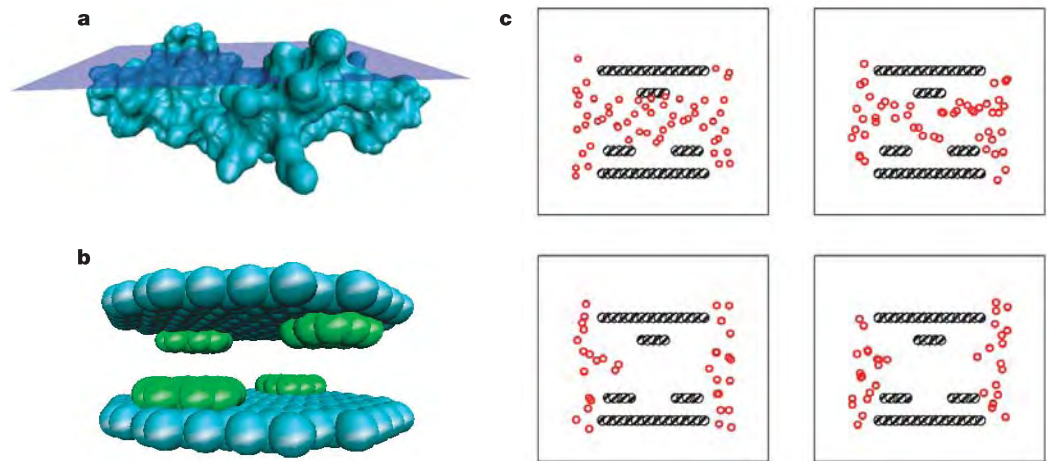


Figure 3 | Dewetting in a model system. **a**, The topology of a melittin dimer. The reference plane shows two bumps in the region between two dimers. **b**, A model system of two graphite-like plates with two argon ‘bumps’ (green balls)

on each plate. **c**, Snapshots of drying transition as a function of time for this model plate system embedded in Simple Point Charge (SPC) water (only oxygen atoms of water are shown; red dots). The inter-plate distance is $6.5 \text{ \AA}$.

Previous work^{3,6} has shown that drying is very sensitive to very small changes in the attractive interactions between hydrophobic particles and water. Our results, showing that drying is exquisitely sensitive to point mutations of Ile to less hydrophobic residues, are consistent with those studies, and demonstrate that in special cases the self-assembly or collapse of proteins can be induced by the evaporation of water (drying) in regions enclosed by their hydrophobic surfaces.

METHODS

The starting structure of the melittin tetramer is taken from the crystal structure deposited in PDB (accession number 2mlt.pdb)¹⁸. The two dimers (chains A and B, and chains C and D) of the tetramer are separated by a distance D , ranging from 4.0 to 8.0 Å, to create a 'nanoscale channel', and are then solvated in water boxes with water molecules at least 8 Å from the protein surface. Twenty Cl⁻ counterions are added to all solvated water boxes to make the system electrically neutral. The final solvated protein systems have up to 25,000 atoms (the actual size varies for different channel sizes D). The OPLSAA force field is used for the protein¹⁹ and the SPC water model for the explicit solvent, with a cut-off of 12 Å for the non-bonded interactions. The GROMACS simulation package is used here for this large system because of its fast speed²⁰. Constant-pressure, constant-temperature (normal pressure and temperature: 1 atm and 300 K) molecular dynamics with a time step of 1.0 fs is used for each system after a conjugate gradient minimization. A total of ten different 'proteins', one native and nine single mutations (each of the three Ile residues mutated to Val, Ala and Gly, respectively), are simulated starting at five different initial separations ($D = 4$ to 8 Å), with up to ten different initial configurations for each separation (same protein conformer but solvated in different water boxes) and up to 10 ns of molecular dynamics simulation, which aggregates to more than 1 μs of total molecular dynamics time. Some of the simulations were run on IBM BlueGene/L development machines.

Received 18 February; accepted 19 June 2005.

1. Ball, P. How to keep dry in water. *Nature* **423**, 25–26 (2003).
2. Wallqvist, A. & Berne, B. J. Computer simulation of hydrophobic hydration forces on stacked plates at short range. *J. Phys. Chem.* **99**, 2893–2899 (1995).
3. Hummer, G., Rasaiah, J. R. & Noworyta, J. P. Water conduction through the hydrophobic channel of a carbon nanotube. *Nature* **414**, 188–190 (2001).
4. Leung, K., Luzar, A. & Bratko, D. Dynamics of capillary drying in water. *Phys. Rev. Lett.* **90**, 065502 (2003).
5. Huang, X., Margulis, C. J. & Berne, B. J. The dewetting and collapse of nanoscale hydrophobic objects. *Proc. Natl Acad. Sci. USA* **100**, 11953–11958 (2003).
6. Huang, X., Zhou, R. & Berne, B. J. Drying and hydrophobic collapse of paraffin plates. *J. Phys. Chem. B* **109**, 3546–3552 (2005).
7. Stillinger, F. H. Structure in aqueous solutions of nonpolar solutes from the standpoint of scaled-particle theory. *J. Solution Chem.* **2**, 141–158 (1973).
8. Lum, K., Chandler, D. & Weeks, J. D. Hydrophobicity at small and large length scales. *J. Phys. Chem. B* **103**, 4570–4577 (1999).
9. Chandler, D. Two faces of water. *Nature* **417**, 491 (2002).
10. Sansom, M. S. P. & Biggin, P. C. Biophysics: water at the nanoscale. *Nature* **414**, 156–159 (2001).
11. Zhou, R., Huang, X., Margulius, C. J. & Berne, B. J. Hydrophobic collapse in multidomain protein folding. *Science* **305**, 1605–1609 (2004).
12. Shea, J., Onuchic, J. & Brooks, C. L. Probing the folding free energy landscape of the src-SH3 protein domain. *Proc. Natl Acad. Sci. USA* **99**, 16064–16068 (2002).
13. Cheung, M. S., Garcia, A. E. & Onuchic, J. Protein folding mediated by solvation: water expulsion and formation of the hydrophobic core occur after the structural collapse. *Proc. Natl Acad. Sci. USA* **99**, 685–690 (2002).
14. Wolde, P. R. T. & Chandler, D. Drying-induced hydrophobic polymer collapse. *Proc. Natl Acad. Sci. USA* **99**, 6539–6543 (2002).
15. Cheng, Y. & Rossky, P. J. Surface topography dependence of biomolecular hydrophobic hydration. *Nature* **392**, 696–699 (1998).
16. Steitz, R. et al. Nanobubbles and their precursor layer at the interface of water against a hydrophobic substrate. *Langmuir* **19**, 2409–2418 (2003).
17. Jensen, T. R. et al. Water in contact with extended hydrophobic surfaces: Direct evidence of weak dewetting. *Phys. Rev. Lett.* **90**, 086101 (2003).
18. Terwilliger, T. C. & Eisenberg, D. The structure of melittin. *J. Biol. Chem.* **257**, 6010–6015 (1982).
19. Jorgensen, W. L., Maxwell, D. & Tirado-Rives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **118**, 11225–11236 (1996).
20. Lindahl, E., Hess, B. & van der Spoel, D. Gromacs 3.0: A package for molecular simulation and trajectory analysis. *J. Mol. Model.* **7**, 306–317 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Castanos and his team for providing the running environment, and M. Eleftheriou, B. Walkup and A. Royyuru for help and support with the BlueGene/L machine. This work was supported in part by an NIH grant and an IBM SUR grant to B.J.B.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.Z. (ruhongz@us.ibm.com) and B.J.B. (berne@chem.columbia.edu).

-  FOCUS
-  SPOTLIGHT
-  RECRUITMENT
-  ANNOUNCEMENTS
-  EVENTS

naturejobs

A boost for the Big Apple

New York is a great place to get a scientific education. But for young researcher looking to switch from academia to local industry, the city has always lagged behind the scientific hubs in California and Massachusetts.

It is true that New York has one of the highest concentrations of world-class universities and hospitals — New York University, Columbia University, Rockefeller University and the Memorial Sloan-Kettering Cancer Center, for example, are all within a short walk or subway ride from each other. But other than a few small incubators, biotechnology companies have largely been absent — especially from Manhattan, where most of the academic research institutions are located.

That may now change, thanks to two developments announced last month that together total more than \$800 million in investment. One will boost the biotech sector, the other, the basic research that spurs it. The first is a \$700-million research complex called the East River Science Park, which will consist of 78,000 square metres of lab and office space. The project, to be built in two phases, will begin construction in 2006; the first 46,000 square metres will be

finished in 2008. It is being funded by the city, and developed by California scientific building specialists Alexandria Real Estate Equities. The company will pay the city \$3 million a year rent, and act as a landlord for the site's tenants.

Meanwhile, New York University is set to receive a donation of some \$105 million from Jan Vilcek, a professor there who made his fortune by co-discovering the blockbuster drug Remicade, which is used to treat Crohn's disease and arthritis. The money will be used to fund basic microbiology research at the university — the kind that led Vilcek to his drug.

Together, the two developments should bring more entrepreneurially minded researchers to the city — and, if Vilcek's generosity is any example, draw in even more funding to fuel the next generation of scientists.



Paul Smaglik, Naturejobs editor

CONTACTS

Publisher: Ben Crowe
Editor: Paul Smaglik
Marketing Manager: David Bowen

US Head Office, New York
 345 Park Avenue South, 10th Floor,
 New York, NY 10010-1707
 Tel: +1 800 989 7718
 Fax: +1 800 989 7103
 e-mail: naturejobs@natureny.com

US Sales Manager/Corporations:
 Peter Bless
 Classified Sales Representatives
 Tel: +1 800 989 7718

**New York/Pennsylvania/
 Latin America:** Kelly Roman
**Midwest USA/Maryland/
 NIH:** Wade Tucker
East USA/Canada:
 Janine Taormina

San Francisco Office
Classified Sales Representative:
 Michaela Bjorkman
 West USA/West Corp. Canada
 225 Bush Street, Suite 1453
 San Francisco, CA 94104
 Tel: +1 415 781 3803
 Fax: +1 415 781 3805
 e-mail: m.bjorkman@naturesf.com

European Head Office, London
 The Macmillan Building,
 4 Crinan Street,
 London N1 9XW, UK
 Tel: +44 (0) 20 7843 4961
 Fax: +44 (0) 20 7843 4996
 e-mail: naturejobs@nature.com

Naturejobs Sales Director: Nevin Bayoumi (4978)
European Sales Manager: Andy Douglas (4975)

Advertising Production Manager: Billie Franklin
 To send materials use London address above.
 Tel: +44 (0) 20 7843 4814
 Fax: +44 (0) 20 7843 4996
 e-mail: naturejobs@nature.com

Naturejobs web development: Tom Hancock
Naturejobs online production: Niamh Shields

European Satellite Office
 Patrick Phelan
 e-mail: p.phelan@nature.com

Japan Head Office, Tokyo
 Chiyoda Building,
 2-37 Ichigayatamachi,
 Shinjuku-ku,
 Tokyo 162-0843
 Tel: +81 3 3267 8751
 Fax: +81 3 3267 8746
Asia-Pacific Sales Director: Rinoko Asami
 e-mail: rasami@naturejpn.com

Expression of interest

Scientists willing to tackle membrane proteins should find opportunities unfolding, says **Hannah Hoag**.

Membrane proteins are fickle. They play essential roles in cellular processes, govern communication between the cell and its surroundings and orchestrate the flux of materials across the cell membrane. They're also the target of about half the drugs on the market. Yet because they are notoriously more difficult to work with than soluble proteins, researchers know little about their structures. The human genome encodes an estimated 10,000 membrane proteins, but less than 1% have had their three-dimensional structures determined.

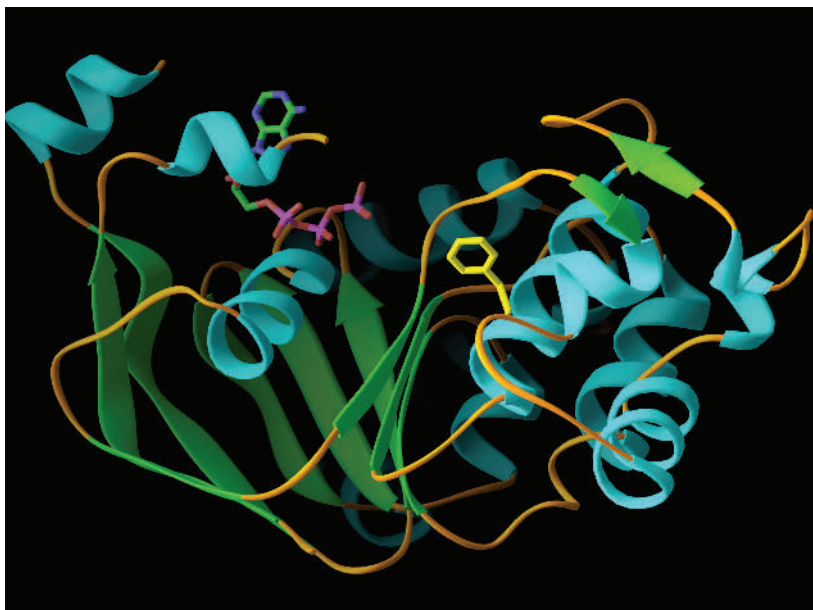
Encouraged by the reams of data produced by recent sequencing efforts, protein scientists now hope that membrane protein structure determination will soon become automated and accelerated. Increases in funding and new technologies may provide enough impetus to launch membrane protein chemistry into lush new territory. Scientists with skills in protein bioinformatics or expression of soluble membrane proteins, and those who can use technology to solve these structures, will be in demand as labs around the world aim to add membrane proteins to international databases over the next five years.

But to do so, they must first identify target proteins and expression vectors able to churn out milligram amounts of easily purified protein. And they'll need to develop solutions that will let techniques such as X-ray crystallography reveal the protein's three-dimensional structure. Then they'll need to use these imaging technologies to take pictures of the proteins, and convert them into three-dimensional images — a slow and expensive process.

Taking the initiative

In 2000, the US National Institutes of Health launched a pilot project called the Protein Structure Initiative (PSI) designed to develop methods and techniques to solve protein structures more quickly and cheaply. Five years and 1,100 structures later, the PSI is entering phase two and going after tougher structures, says PSI director John Norvell. Labs in Canada, Japan and Europe are following suit, intensifying their efforts to solve the structures of membrane proteins in both eukaryotes, whose cells have a nucleus, and prokaryotes, whose cells do not. Some groups will develop technologies; others will focus on solving structures.

But the transition from soluble proteins to membrane proteins is full of challenges. Initially, it will need scientists who can identify targets: those that are representative of large families or that are biologically significant. People skilled in protein bioinformatics will be in high demand and will set the tone for the rest of the work. These techniques are likely to be very different from the ones used for the smaller soluble proteins, says Wayne Hendrickson, director of the New



LEWIS ET AL

The three-dimensional structures of membrane proteins are difficult to determine.

York Consortium on Membrane Protein Structure (NYCOMPS) based at the New York Structural Biology Center.

"If we do the bioinformatics right, one structure for a family will inform us greatly about the structures of all the family," says Hendrickson. Although it hasn't started recruiting officially, NYCOMPS will be building a protein-production facility from scratch, and recruitment for this site will dominate its short-term planning.

Expression remains the most significant bottleneck in the process. G-protein-coupled receptors (GPCRs) are one type of membrane protein that interests researchers, but they are difficult to produce. These molecules make up the largest known protein family and play a key role in how other proteins in the cell interact. They are the target for about 30% of today's pharmaceuticals because of this role and because lots of small molecules bind to them.

"Until we can produce large amounts of pure, properly folded human ion channels or human GPCRs, there will be limited progress on structural studies of

FRINGE BENEFITS

Postdoctoral researchers Richard Darby and Mohammed Jamshad are motivated by the challenges they face in their work. But they're also enjoying the fringe benefits that come from joining a large consortium such as the European Membrane Protein Consortium.

"The whole project fitted the bill in terms of what I wanted," says Jamshad, a microbiologist. "The research was going to be challenging and there would be plenty of opportunities for training."

Both Darby and Jamshad travelled to Chalmers University in Gothenburg, Sweden, early in their projects to hone their medium-scale fermentation skills — a method that increases protein yields from *Pichia pastoris* yeast.

"A large consortium such as this opens so many doors and benefits," says Darby, a molecular biologist. "We've had the opportunity to train in other areas where we do not have the expertise or where we have limited expertise."

H.H.

them,” says Stephen Burley, chief scientific officer of Structural GenomiX in San Diego, California.

Part of the problem is that researchers still need to identify expression systems that can produce the relatively large quantities of the notoriously unstable proteins necessary for imaging. Few believe that a single technique will do the trick. The European Membrane Protein Consortium (E-MeP) is one of many groups that has chosen several expression systems — microbial, mammalian and cell-free — and will systematically determine the best for a given protein. “There will be no panacea,” says E-MeP coordinator Roslyn Bill, of Aston University in Birmingham, UK.

Concerted effort

E-MeP, a €10.4-million (US\$12.7-million) European Commission programme, is a large consortium of labs in Britain, France, Germany, Sweden and the Netherlands. Bill says it hopes to get 20 new structures from the 300 target membrane proteins with biological or industrial interest during its five years of funding. Scientists who can develop new cloning and reproduction methods and are comfortable working with eukaryotic and prokaryotic expression systems will excel in their job search, she says.

Two of the PSI's new bases — the Center for Structures of Membrane Proteins at the University of California, San Francisco, and NYCOMPS — will be dedicated to membrane protein structural biology. Both will develop technologies and will test expression systems and detergents, as well as experimenting with other modifications, such as antibodies, to increase the solubility of proteins for crystallization. The centres will need people with specialist knowledge of detergents and an aptitude in protein purification.

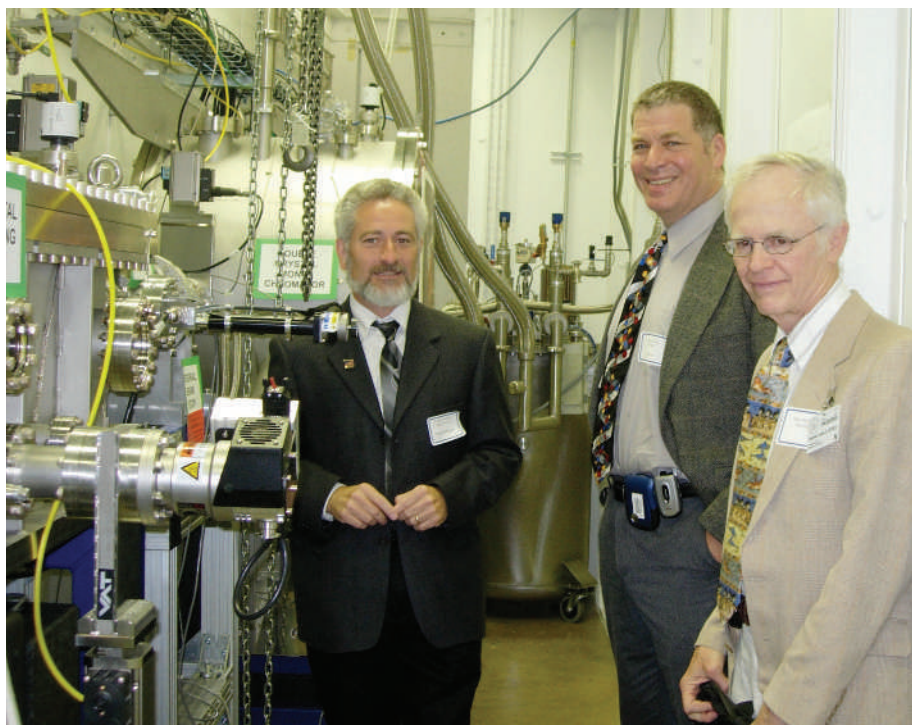
Aled Edwards, executive director of the not-for-profit Structural Genomics Consortium (SGC), has room for people with a flair for protein purification. “We’re always recruiting,” says Edwards.

The SGC has a three-year mandate to put 350 three-dimensional structures into an unrestricted database (see *Nature* **435**, 547; 2005). It is based at the universities of Toronto and Oxford and the Karolinska Institute in Stockholm, with its membrane protein efforts centred in Britain. Rather than focus on technologies, the SGC will target proteins important in certain diseases including diabetes and cancer.

Joining a consortium can offer younger scientists the opportunity to expand the breadth of their knowledge. “We’re trying to train scientists so that they have an appreciation of all the areas that are required to be a good membrane protein structural biologist,” says Bill. E-MeP set aside money from its budget for professional development (see ‘Fringe benefits’, opposite).

Industrial angles

Those with a knack for crystallization might look to industry. Structural GenomiX specializes in structure-based drug discovery, with a primary focus on oncology products. The company uses high-throughput X-ray crystallography to screen the binding activity of drug candidate molecules to target proteins. In July, the company expanded its collaboration with the Cystic Fibrosis Foundation to begin doing drug discovery against mutant forms of a protein involved in the disease. Structural GenomiX had 20 positions — ranging from bioinformatics to protein expression to



John Norvell (right) examines new laser beam lines for structure determination.



“We’re trying to give scientists an appreciation of all the areas needed to be a good membrane protein structural biologist.”

— Roslyn Bill

crystallization — posted on its website in August. The company will hire about a dozen people to support the work, which is being carried out at the New York Structural Genomics Research Consortium.

Harren Jhoti, chief scientific officer at Astex Therapeutics in Cambridge, UK, says that finding people with the right mix of biophysics and X-ray crystallography can be difficult. Rather than identify potential drug targets, one of the company’s initiatives hinges on the protein cytochrome P450, which is involved in drug metabolism.

Many academic team leaders fear that the best scientists will head to industry for higher salaries. “It is extremely difficult to get good people who want to work in an academic environment,” says Larry Miercke, a professional associate at the Membrane Protein Expression Center at the University of California, San Francisco. But the opportunity to publish in high-impact journals and to be involved in the full breadth of research dissuades many from leaving.

The ability to switch gears and make the leap from academia to industry is more of a management challenge than a technological one, says Burley, who gave up an endowed professorship at Rockefeller University in 2002 to join Structural GenomiX and do high-throughput crystallography. “The science, the technology is all the same.” Deadlines are tight and finances are limited. But, above all, researchers must be able to produce a result on a given day. New recruits must recognize that being able to work with others in pursuit of a common goal is a prized quality in industry.

Laboratories around the world are increasing their efforts to solve the crystal structures of membrane proteins. And scientists who are willing to tackle these fickle molecules should flourish in both industry and academia because right now their skills are in demand. ■

Hannah Hoag is a journalist based in Montreal, Canada.

WEB LINKS

Structural Genomics Consortium
 ♦ www.sgc.utoronto.ca
 Protein Structure Initiative
 ♦ www.nigms.nih.gov/psi
 European Membrane Protein Consortium
 ♦ www.e-mep.org
 New York Structural Genomics Research Consortium
 ♦ www.nysgrc.org
 Membrane Protein Expression Center
 ♦ mpec.ucsf.edu/index.htm
 Structural GenomiX
 ♦ www.stromix.com
 Astex Therapeutics
 ♦ www.astex-therapeutics.com

Nostalgia

A novel resurrection.

Hiromi Goto

Mercury Lam Meinhart's thudding heart almost burst through her chest. There's no way for them to detect the theft, she reminded herself. Because what she had taken was organic. A single strand of hair with its root intact. If she kept her wits she'd leave this mausoleum unnoticed and be on a life trajectory far beyond the stunning boredom of an under-published literary historian. To be free from eating Soygen 3 for perpetuity! This thought alone brought tears of joy to her eyes.

The soft noise of her polyform shoes. Desperation seeped from her second-best suit. Her thudding heart. Dimly, she wondered how she'd been capable of this act. Weren't the mandatory gene tweaks meant to eliminate all abhorrent behaviour?

"Miss —"

Mercury felt faint.

"Miss, Miss!"

Mercury, partially digested Soygen rising in her gorge, turned around.

"Ha-haaaa!" a young man crowed. His companion, a young woman with jewel lights in her scalp, tried to hop out the patterns of their game, but she stumbled.

"Miss, again!" the young man gloated.

A micro siren wailed. "Game-playing is forbidden in the museum," a quiet digitalized voice declared. "Cease activity. You have five minutes to exit the facility. Non-compliance will result in loss of leisure credits."

"Delete!" the young woman cursed. "This was a requirement for Victorian Lit. class!"

"Come on," the young man mumbled. "We can go on a sim-tour instead."

As Mercury turned around, her heart slowly began beating again. A smile played on her lips.

"Whaddya got?" Leo Yoshida slurred.

Image and sound were not to grade, but Mercury was certain her ex was in the middle of a sim-high. What time was it in Hong Kong? She was sure Leo didn't care. "Remember my pitch? A reality show, but high art?"

"Oh, yah. Nostalgia angle. We're ripe for it here, Merc. It could fly. You always had good ideas."

"More than an idea. I've already begun



ripening the perfect candidate. Guess who?"

"Shakespeare? Uhhh, the *Crime and Punishment* guy? I dunno."

"Is your barrier secure, Leo?"

"You know I'm always clean," he leered.

"My old man made us split up because of the merger, but I've always had a soft spot for you."

Mercury rolled her eyes, but she couldn't stop smiling. She could see her credit ratings breaking through the 'class-free' barrier. "Brontë," she whispered. "I got a Brontë!"

"Who'zat?" Leo asked.

"Gates!" Mercury swore. "Didn't you attend Required Lit. lectures? I'm cloning Charlotte or Emily Brontë. One of the sisters. I took a hair from a mourning brooch at their shrine in Yorkshire."

"Whose facilities are you using?" Leo asked, suddenly all business. He must have turned off his sim program. "What's your estimated time of fruition?"

"I've rented an off-Net portable speed-Queen. ETF in six months."

"I can't believe they're still legal out there! They've lost their licence in Asia Major and Australasia," Leo spluttered.

"Tried and true," Mercury wrinkled her nose. "You know their motto. 'No one need ever outlive their pet again!' Listen, I've sent across the specs already. Find four more writers to make this show work. Otherwise, I'm walking with my Brontë. I can groom her as a novelty personality."

"No! I'll have a contract for you in a few

minutes. How will I find the other nostalgia writers?" Leo moaned.

"Include me as a consultant. Ideas are my forte," Mercury bared her teeth.

All those writers who had left pieces of themselves. Archives and special collections. Treasure hordes, they were. Filled with papers with pieces of skin, hair, even blood. A splatter of Hemingway. A drop from Mishima's seppuku.

And not only writers! The second season could be tortured artists. Van Gogh. Frida Kahlo. They could be groomed in special holo environs to replicate their original circumstances. Their 'development' could be broadcast on the Nets and betting pools could be arranged. The copies still weren't exact replicas and the technique hadn't eradicated every dysregulation, but a few abnormal traits would only make the artists

more interesting. More tortured. It would be the biggest credit-making artistic freak show in the history of entertainment.

"I've sent the contract!" Leo's eyes shone. "Merc, Merc, what can I have delivered so you can start celebrating now?"

"Mmmm," Mercury closed her eyes. This was what it felt like to 'have the splice of life'. Oh so heady... "I've always wondered what hydroponic oysters tasted like."

There was no denying it. The Brontë was a male.

"Gates! Gates!" Mercury swept the data from her desk. She couldn't risk going to Yorkshire to steal another piece of their lives. But Branwell Brontë! He'd had been a second-rate artist. A clichéd alcoholic.

Mercury tapped her finger against her lips. No need to throw out the clone with the amniotic fluid.

It wasn't too late to introduce hormone therapy. Branwell could be turned into a female. And Emily had been an odd creature, practically autistic.

Mercury smiled. No one would notice a thing. And by the time they did, they'd be well into their second season with a whole new cast.

Mmmmmmm. She loved oysters. ■

Award-winning author of *Chorus of Mushrooms* and *The Kappa Child*, Goto is a genre-bending writer whose most recent book is a collection of short stories called *Hopeful Monsters*. She has a novel pending with Penguin Canada.

JACEY

Abstractions



FIRST AUTHOR

The paper on page 88 is a landmark for Ze 'Ginger' Cheng: she has never before been a 'first author' in *Nature*. A computer programmer and

database administrator at the University of Washington's Department of Genome Sciences, Cheng spent most of the past year furiously writing computer code. The fruit of her labour is an analysis of DNA duplication differences between human and chimp.

Cheng was looking for duplications of DNA that appear in both genomes, as well as those that are unique to either species. She was also seeking duplications with a different number of copies in humans and chimps. Such duplications are thought to be a force for evolutionary change in genomes and are sometimes related to human diseases. *Nature* caught up with Cheng to find out more about her work.

How massive was this project?

It was big. I worked eight or nine hours a day on this for a year. I helped analyse around 100 gigabytes of data. For one table, I looked at 700 files. The project has been the dominant thing in my life for a year. It has *been* my life for a year.

What sort of challenges did you face?

We had a very heated lab meeting about the critical thresholds to detect recent duplications using data from two different species.

Why was this a challenge for you?

My background is in cell biology and molecular biology. Genomics research is very different. I had to read some papers to learn the concepts and to get used to what everybody was talking about.

What did the other members of the group do?

Evan Eichler came up with all the analysis for the paper and set out the way we should approach the topic. Eray Tuzun performed the analysis for duplication in the chimp genome assembly. Xinwei She did the gene-expression analysis of duplicate genes using data from our collaborator, Svante Pääbo.

Any critical points?

There was one day when I was very ill, but I had to finish what I was working on. I like my job, but it involves pressures. If you don't conquer those pressures, you don't accomplish anything.

How did you deal with the stress?

I jog to shed the pressure. When I get mentally exhausted, I want to be physically exhausted to relax.

How do you feel about the paper now?

I was not very clear of the big picture at the beginning. But when I look back, it is so beautiful. ■

MAKING THE PAPER

Sally McBrearty and
Nina Jablonski

How a trip to Kenya put the modern chimpanzee on the map.

Examining a fossilized molar in Nairobi last October, Nina Jablonski was thinking just one thing: "This is no monkey." It should have been. After all, Jablonski, an anthropologist at the California Academy of Sciences, was in the Kenyan capital to study a collection of monkey fossils unearthed at the Kaphthurin Formation in the East African Rift Valley.

But the size and shape of the tooth were wrong. "My first thought was, 'Oh my gosh, this is a gibbon,'" Jablonski says. "Then I realized, no, it was almost certainly a chimp." Further searching among the monkey fossils turned up a second anomaly; this time, an incisor.

The collection had been amassed by Sally McBrearty, an anthropologist at the University of Connecticut, and her team who were in Kenya looking for stone tools and fossils. Jablonski's musings echoed McBrearty's suspicions. "I believed it was an ape," McBrearty recalls. The result was the paper co-authored by the pair on page 105 of this issue.

Jablonski compared the fossils with teeth of modern chimps at the National Museums of Kenya. The molar, with its distinctive crown features, was a perfect match. "I sent off a very excited e-mail to Sally," Jablonski says. "She was thrilled because she had suspected that the teeth did not belong to a monkey."

McBrearty immediately sent her team back into the field. By March, the researchers had found more chimp teeth, bringing the total to four, three of which are described in this issue.

"This is the first unequivocal example of a modern chimpanzee in the fossil record — no question," Jablonski says.

McBrearty's team has yet to uncover any other chimp remains. Teeth are often the only fossils left, as they are harder than other parts



Sally McBrearty (left) and Nina Jablonski.

of the body and less likely to be crushed or decay. "I'm hoping to go back and spend more time looking for the rest of this creature," says McBrearty.

The discovery of the chimp teeth was a major surprise, Jablonski says. No fossils of modern apes had ever been found in Africa — and to cap it all, the remains came from the

Rift Valley. Today, chimps are found only in western and central Africa; the Rift Valley was seen as a geographical barrier to the species, preventing its spread farther east. But the teeth, which are about 500,000 years old, put chimps in a part of the continent

where they weren't known to exist. And, more importantly, they put the chimps side-by-side with a hominid.

Although she is aware of implications for evolution and anthropology, Jablonski sees some irony in the discovery of the teeth. She has been working on monkey fossils for 30 years, and has made many finds that characterize that group's evolution, but few people outside her speciality "gave a hoot or a holler" about such work. The chimp teeth attracted much more attention because of the close evolutionary link between the apes and humans.

Jablonski has now returned to her primary research interest. "I'm happily going back to monkeys," she says. "But I certainly won't ignore any chimps that cross my trail." ■

"I'm going back to monkeys — but I certainly won't ignore any chimps that cross my trail."
— Nina Jablonski

QUANTIFIED GERMANY

A numerical perspective on *Nature* authors.

Several authors in Germany present research in *Nature* this week. Andreas Richter and his team, working in Bremen and Hamburg, use satellite data to measure levels of nitrogen dioxide in the lower atmosphere and report a worrying increase in nitrogen oxide emissions over industrial areas of China (see page 129).

Thomas Blankenstein and Gerald Willmsky, at the Charité in Berlin, use mice to look at cancer immunology and show how sometimes certain types of tumour can evade an immune response (see page 141).

And Svante Pääbo, based in Leipzig, is a member of the team reporting results of a genome-wide comparison of DNA similarities between humans and chimpanzees (see page 88).

537 submissions to *Nature* have come from Germany since 1 January 2005 (total global submissions = 8,987)

269 authors published in *Nature* so far this year were working in Germany (total number of published authors = 3,725)

22 authors working in Germany have had more than one paper published in *Nature* since 1 January 2005.

8 authors working in Germany are presenting original research in *Nature* this week.